UNIVERSITY OF LONDON THESIS

Degree PhD Year 2007 Name of Author DIAS DE CASTRO RODRIGUES, Artur Filipe

Structural Studies on the H3 Influenza A Virus Haemagglutinin

Receptor Binding and Membrane Fusion

Artur Filipe Dias de Castro Rodrigues

A thesis submitted in fulfilment of the requirements of University College London for the degree of Doctor of Philosophy

May 2007

Division of Virology
Division of Molecular Structure
MRC-National Institute for Medical Research
The Ridgeway
Mill Hill
London

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

In this project, the receptor binding and membrane fusion activities of the membrane-anchored glycoprotein haemagglutinin (HA), H3 subtype, of the influenza A virus were studied.

Influenza viruses from different hosts can distinguish between α2-3 and α2-6 linkages that sialic acid (Sia) forms with the penultimate saccharide residue of the receptor. Human viruses prefer α2-6 linkages and avian viruses α2-3 linkages. The side chain of residue 226 of the HA receptor binding site (RBS) is involved in the specific recognition of those linkages. The HA of the 1968 Hong Kong (HK) pandemic virus (H3 subtype), contained in the recombinant X-31 virus, has a 226-leucine and prefers binding to α2-6 linkages. The L226Q HA of the variant virus X-31/horse serum (X-31/HS), corresponding to a single-site mutant of the X-31 HA having 226-glutamine, prefers α2-3 linkages. To define the molecular interactions of the L226Q HA with the virus receptor, the crystal structures of L226Q HA in complex with the receptor analogues lactoseries tetrasaccharide a (LSTa), α2-3-terminated, and lactoseries tetrasaccharide c (LSTc), α2-6-terminated, were determined. The structures show the saccharide residues Sia and galactose (Gal) of LSTa and only the Sia of LSTc bound on the L226Q HA RBS, which correlates with the higher affinity of L226Q HA for the α2-3 linkage. However, the L226Q HA binds both the *trans* and *cis* configurations of the Sia-Gal glycosidic bond of LSTa, which has never been observed before. These results are discussed in comparison with data from binding assays and the available crystal structures of the X-31 HA, the H3 avian HA of the influenza virus A/duck/Ukraine/63, a potential precursor of the 1968 HK pandemic virus, and other HAs, of different subtypes and hosts of origin, in complex with LSTa and LSTc.

Upon endocytosis, at the low endosomal pH, the HA undergoes an irreversible conformational change associated with the fusion of the viral and endosomal membranes, a

process by which the virus enters the target cell. To gain further insights into the membrane fusion mechanism catalysed by the influenza HA, the purification and crystallisation trials of three different X-31 HA forms representative of the neutral pH and fusion pH-induced conformations, containing the membrane-interacting segments fusion peptide and transmembrane anchor, were carried out with the future aim of their crystal structure determination. The procedures for the purification and crystallisation trials of the X-31 HA proteins in different conformations are discussed in the context of the general purification and crystallisation of membrane proteins. No promising hint has yet been obtained in the crystallisation trials.

I, Artur Filipe Dias de Castro Rodrigues, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Artur Filipe Dias de Castro Rodrigues carried out his doctoral research at the National Institute for Medical Research (NIMR), in the Division of Virology and Division of Molecular Structure under the supervision of Sir John Skehel and Prof. Guy Dodson, and was funded, as a student of the Gulbenkian Ph.D. Program in Biomedicine, by the Calouste Gulbenkian Foundation and the Science and Technology Foundation (Portugal), and subsequently by the Medical Research Council (MRC).

Acknowledgements

I would like to thank my principal supervisor, John Skehel, for giving me the opportunity to carry out this project in his lab, his support, guidance and openness to discussions, and for funding me.

I am particularly indebted to Steve Wharton, for all his support, advice and endless availability for discussions concerning practically every aspect of this project, and comments on this text. I am thankful to David Stevens for his countless advice on the purification of haemagglutinin. I am also thankful to both for general supply of reagents and materials. I would like to thank Rose Gonsalves and David Stevens for teaching me how to grow viruses in chicken eggs and purify them, and giving me some purified virus samples, and Alan Douglas for his technical advice and supply of antibodies and Barry Ely for general advice. I am grateful to Lesley Calder for carrying out an electron microscopy analysis. A big thanks to the everyone in the influenza lab, John Skehel's group, for the uplifting mood, particularly Barry for all the good humour. I am also grateful to Mikhail Matrosovich for discussions about receptor binding by influenza haemagglutinin.

I would like to thank my second supervisor, Guy Dodson, for the warm reception to the Division of Molecular Structure, encouraging comments, advice on data analysis and the writing of this text, and discussions about X-ray crystallography. I am very grateful to Steve Gamblin for all the help with crystal harvesting and X-ray diffraction data collection, processing and refinement, and comments and suggestions on this text. I would like to thank Steve Smerdon for general advice on X-ray crystallography and comments on this text. I am very grateful to Lesley Haire for all the advice and discussions about crystallisation and insightful suggestions. Also, my thanks to Rupert Russell for his help with initial analysis of crystals and final structure refinement, and Junfeng Liu for the help with data processing and image representation. I am also thankful to Lesley, Katharina

Berchner and Nishi Vasisht for help with automatic setting up of drops. I would like to thank Steven Howell for carrying out the peptide mass fingerprinting, and corresponding data base search, and advice on the required sample preparation. I am grateful to Phil Walker, for all the help with general computer and X-ray diffractometer utilization. My thanks go also to Neil, Barbara, Julie and Leon, for suggestions on data processing and refinement.

I am thankful to everybody in the Divisions of Virology and Molecular Structure for the friendly atmosphere, which has helped me to overcome the inherent difficulties to one's scientific path.

I would like to thank all the friends I made at NIMR for making my time spent in the institute an enjoyable experience. We shared a lot of good moments, not only on the day-by-day basis at the institute but also going out to explore London, visiting some beautiful parts of the UK, particularly with the fantastic team of the NIMR's Hill Walking Club, and travelling abroad to a couple of exciting destinations.

I am grateful to my sponsors in Portugal, the Calouste Gulbenkian Foundation and the Science and Technology Foundation, for the award of a Ph.D. studentship and funds to cover for tuition and bench fees.

I would like to thank all my friends in all different places for their support throughout my Ph.D..

I will be forever in debt to my family for all their love.

Contents

List of Figures

List of Tables

List of Abbreviations

BHA	Bromelain-released HA
CAM	Chorio-allantoic membrane
$C_{12}DAO$	Lauryldimethylamine oxide
$C_{12}E_{10}$	Polyoxyethylene 10 Lauryl ether
CMC	Critical micellar concentration
DMP	Dimethyl pimelimidate
DTT	Dithiothreitol
$EBHA_2$	*E. coli*-expressed $TBHA_2$
E. coli	*Escherichia coli*
EHA_2	*E. coli*-expressed soluble HA_2
EM	Electron microscopy
ER	Endoplasmic reticulum
F185	*E. coli*-expressed flag-tagged HA_2
Gal	Galactose
GalNac	N-acetyl-D-galactosamine
Glc	Glucose
GlcNac	N-acetyl-D-glucosamine
HA	Haemagglutinin
HA_0	HA precursor
HDVD	Hanging drop vapour diffusion
HK	Hong Kong
HS	Horse serum

IE	Ion exchange
kDa	Kilo-Dalton
kV	Kilo-Volt
LST	Lactoseries tetrasaccharide
M	Matrix protein
mAb	Monoclonal antibody
MALDI	Matrix assisted laser desorption/ionisation
MB	Microbatch
MDCK	Madin-Darby canine kidney cell
2-ME	2-mercaptoethanol
MME	Monomethylether
MW	Molecular weight
MWCO	Molecular weight cut-off
NA	Neuraminidase
NCBI	National Center for Biotechnology Information
NCS	Non-crystallographic symmetry
Neu5Ac	N-acetylneuraminic acid
Neu5Gc	N-glycolylneuraminic acid
NMR	Nuclear magnetic resonance
NP	Nucleoprotein
NR-LBD	Nuclear Receptor – Ligand Binding Domain
NS	Non-structural protein
OG	Octyl-β-D-glucopyranoside
ON	Overnight
PA	Polymerase acid

PB	Polymerase basic
PBS	Phosphate-buffered saline
PDB	Protein Data Bank
PEG	Polyethylene glycol
RBS	Receptor binding site
RCSB	Research Collaboratory for Structural Bioinformatics
RNA	Ribonucleic acid
RT	Room temperature
SDG	Sucrose density gradient
SDS-PAGE	Sodium dodecyl sulfate – polyacrylamide gel electrophoresis
Sia	Sialic acid
SL	Sialyllactose
SLN	Sialyllactosamine
SN	Supernatant
*ss*RNA	Single-stranded RNA
$TBHA_2$	Thermolysin-solubilised BHA_2
TI	Trypsin inhibitor
Tris	tris[hydroxy methyl] aminomethane
USF	Uppsala Software Factory
*v*RNA	Viral RNA
*v*RNP	Viral Ribonucleoprotein

1. Introduction

The receptor binding and membrane fusion activities of the HA glycoprotein of influenza virus A play a major role at the initial stages of infection. The X-ray crystallographic studies of HA in complex with receptor analogues, and low pH-treated HA fragments are important for a better understanding of the virus specificity and the mechanisms it uses to enter the target cell.

1.1 Initial Isolation of the Influenza Virus

The influenza virus as a causative agent of human disease was not isolated until 1933 at the MRC-NIMR, from the filtered throat washings of experimentally infected ferrets (Smith et al., 1933). The animals developed influenza-like illness, transmitted the disease to other ferrets and were rendered immune to re-infection. The first virus isolated from humans, presumably was accidentally transmitted by an infected ferret. The virus isolated from the throat washings of the infected human proved to be mouse-adapted as was that given to the ferret (Smith and Stuart-Harris, 1936).

1.2 Classification of Influenza Viruses

Influenza viruses belong to the *Orthomyxoviridae* family of enveloped viruses, having a segmented genome of single-stranded ribonucleic acid of negative polarity [*ss*RNA*(-)*]. They are classified into three genera - influenza A, B and C - based on differences of their nucleoprotein (NP) and matrix 1 (M1) protein antigens (WHO, 1933; Schild, 1972). Influenza A and B contain eight genome segments, whereas influenza C contains seven segments. Influenza A viruses infect various mammals and birds, whereas influenza B and C have only been isolated from humans, with the exception of some type C isolates from pigs (Guo et al., 1983) and a type B isolate from seals (Osterhaus et al., 2000). Influenza A

viruses are subclassified into subtypes, which are distinguished by the antigenicity of the HA and neuraminidase (NA) envelope glycoproteins. To the present time, sixteen HA (H1-H16) and nine NA (N1-N9) subtypes have been identified (Schild et al., 1980; Fouchier et al, 2005). All sixteen HA subtypes circulate in aquatic birds. Of these, only three have evolved in humans (H1, H2, H3) causing pandemics, at least two in pigs (H1, H3) and two in horses (H3, H7) (Webster et al., 1992).

1.3 The Influenza A Virion

1.3.1 Virion composition

The eight *ss*RNA segments (McGeoch et al., 1976) of an influenza A virus particle (Figure 1A) code for ten viral proteins. Each of the RNA segments is encapsidated with NP, forming a coiled hairpin structure (Pons et al., 1969; Compans et al., 1972; Jennings et al., 1983) with the subunit proteins of the polymerase complex, polymerase basic 1 (PB1), polymerase basic 2 (PB2) and polymerase acid (PA) attached to the 5' and 3' ends of the RNA (Hagen et al., 1994). This assembly of viral RNA (*v*RNA), NP and the polymerase proteins forms the viral ribonucleoprotein (*v*RNP) particle. The influenza virion is enveloped by a host-derived lipid bilayer (Kates et al., 1961) containing the membrane-anchored glycoproteins HA and NA, being HA the more abundant: a virion usually contains 300-400 HA and approximately only 50 NA molecules (Ruigrok et al., 1984; 1998). Only a few copies per virion (14-68) of a transmembrane protein, the ion-channel matrix 2 (M2) protein, are embedded in the viral envelope (Zebedee and Lamb, 1988) as tetramers (Sugrue and Hay, 1991). A shell of M1, the major structural component of the virus particle, underlies the viral membrane (Schulze 1970; 1972). The non-structural (NS) proteins NS1 and NS2 have multiple post-transcriptional functions, and occasionally the NS2 protein is packaged into virus particles (reviewed in Lamb and Krug, 1996).

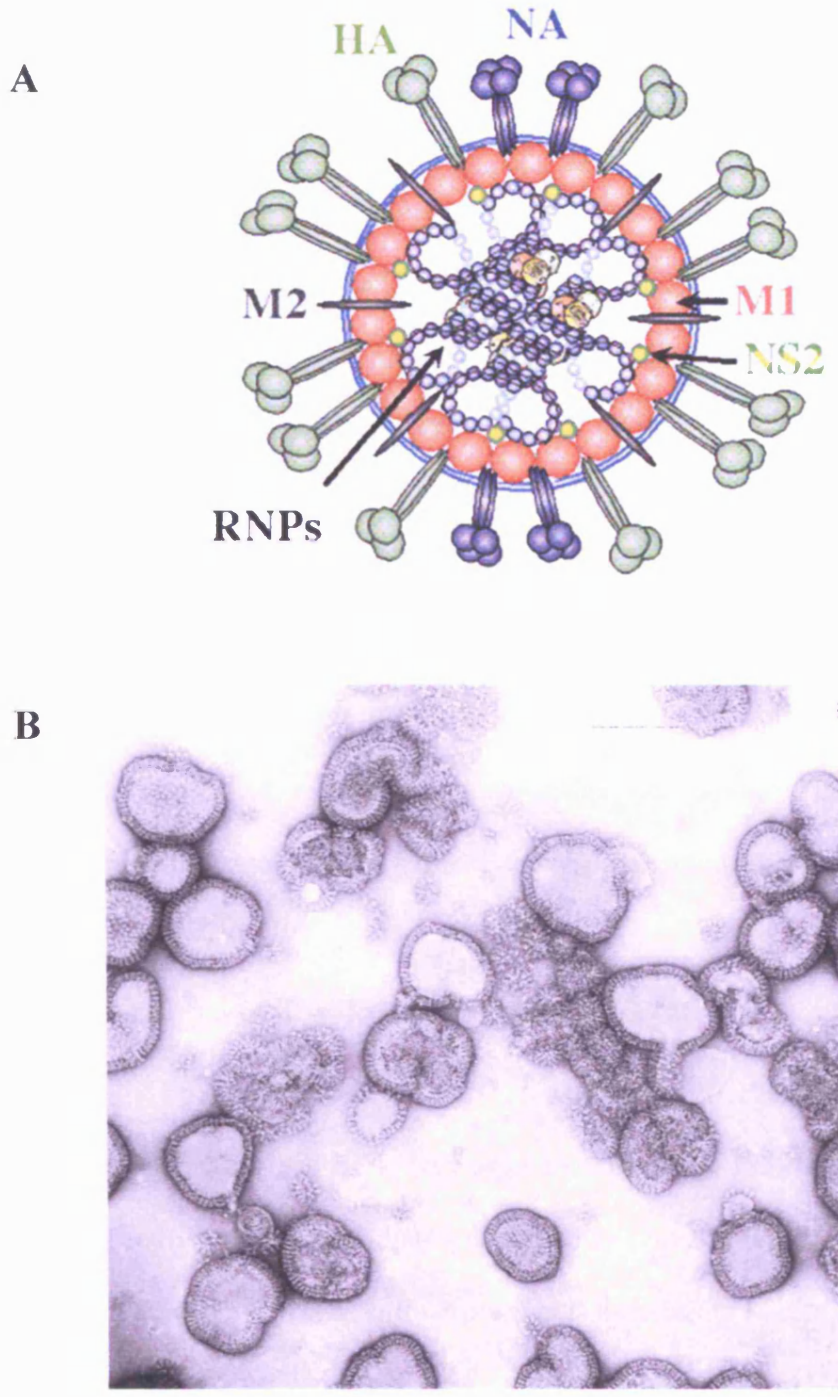


Figure 1. Composition and morphology of influenza A viruses

(A) The influenza virion envelope contains the membrane-anchored glycoproteins HA and NA, and the ion channel M2 protein. The viral envelope is lined on the interior face by the M1 protein, which encloses the *v*RNPs. Each of the latter includes one of the eight genomic *v*RNA segments (see text for details). [Adapted from: http://www-micro.msb.le.ac.uk/3035/Orthomyxoviruses.html].

(B) Negative-staining electron micrograph of influenza virus X-31 (H3N2) virus particles. The picture was taken by L. J. Calder (Division of Virology, MRC-NIMR, London).

1.3.2 Virion morphology

Electron microscopy (EM) studies have shown that influenza virions are structurally diverse, ranging from small spherical particles of 80-120 nm in diameter (Taylor et al., 1943) to long filaments with a length of several μm (Mosley and Wyckoff, 1946). The virion surface glycoproteins HA and NA are seen as spikes protruding from the virus envelope. A progressive change from filamentous to spherical morphology of clinical isolates of human and animal viruses, which are predominantly filamentous (Chu et al., 1949), has been observed upon serial passages in embryonated hens' eggs (Burnet and Lind, 1957; Choppin et al., 1960). Other morphologies seem to be associated with specific techniques of sample preparation (Nermut and Frank, 1971; Ruigrok and Hewat, 1991). An electron micrograph of X-31 (H3N2) virions, used in this study (see *2.1.1*), is shown in Figure 1B. These viruses are spherical and other morphologies are linked with certain conditions of sample preparation (L. J. Calder, personal communication).

1.3 The Influenza Virus Replication Cycle

Infection is initiated by attachment of influenza viruses to target cells, by interaction of the HA with sialylated receptors (see *1.6.1*) (reviewed in Gottschalk, 1959). The virus particles are then internalised by receptor-mediated endocytosis (Fazekas de St. Groth, 1948). A schematic of the replication cycle of influenza virus is depicted in Figure 2. The low pH of the endosome triggers an irreversible conformational change in the HA molecule which mediates the fusion of the viral and endosomal membranes (White et al., 1981; Skehel et al., 1982; Bullough et al., 1994). Just before the membrane fusion process, the interior of the virion is acidified by the proton transport activity of the ion-channel M2 (Pinto et al., 1992; Bui et al., 1996; Chizhmakov et al., 1996), leading to dissociation of M1 from the *v*RNPs, which is necessary for the entry of the viral genome into the nucleus (Bukrinskaya

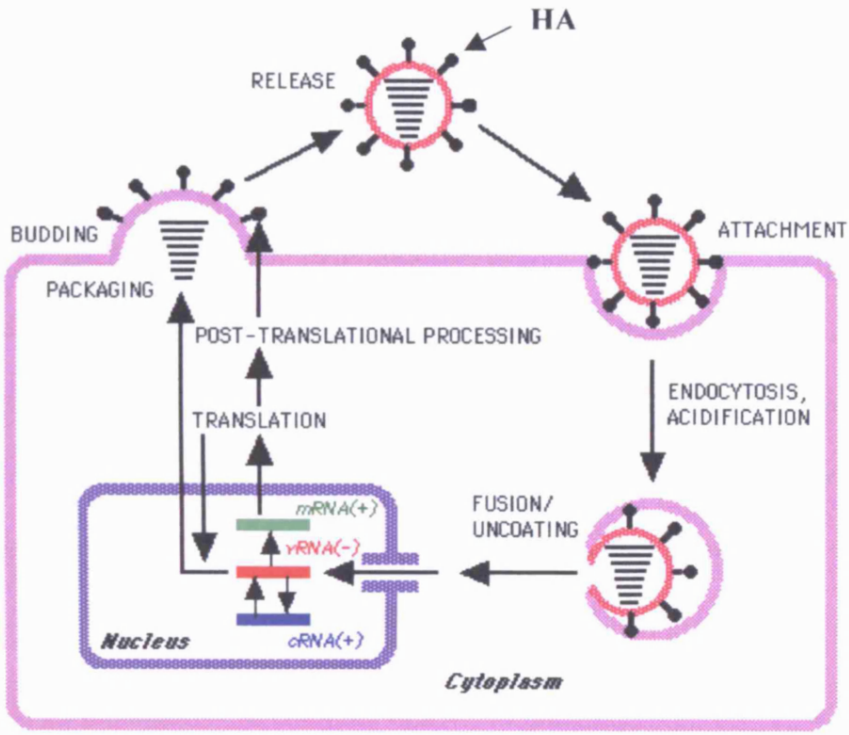


Figure 2. The replication cycle of influenza A virus

Simplified schematic showing the major steps of the influenza virus A replication cycle.

The virus attaches to the target cell membrane via the interaction of the viral envelope HA spikes to cell surface Sia-terminated oligosaccharides. After endocytic internalisation and acidification of the endosomal vesicle, the HA mediates the fusion between viral and endosomal membranes and the M2 proton channel acidifies the interior of the virion leading to disassembling of the M1 protein and release of *v*RNPs into the host cell cytoplasm. Once in the nucleus, the segmented viral genome is replicated and transcribed. After co- and post-translation modifications the viral proteins assemble at the cell surface, and the genome segments are packaged into newly formed virus particles. These are released from cell surface receptors by the activity of the viral membrane NA protein, which cleaves off terminal Sia sugar residues from oligosaccharides (see text for details). This NA activity also prevents the aggregation of virus particles.

[Adapted from: http://www-micro.msb.le.ac.uk/3035/Orthomyxoviruses.html]

et al., 1982; Martin and Helenius, 1991). Once in the nucleus, the *v*RNA*(-)* is transcribed into messenger RNA [*m*RNA*(+)*] and replicated into the complementary RNA [*c*RNA*(+)*] by the proteins PA, PB1 and PB2 of the polymerase complex (Hay et al., 1977; Braam et al., 1983; Honda et al., 2002). The NP is required, together with the polymerase complex proteins, for the synthesis of the *v*RNA (Beaton and Krug, 1986; Portela and Digard, 2002). The envelope proteins HA, NA and M2 are synthesised by ribosomes associated with the endoplasmic reticulum (ER) (Compans, 1973a; Hay, 1974; Lamb et al., 1985). After insertion in the ER membrane (McCauley et al., 1979; Bos et al., 1984; Hull et al., 1988), their folding, oligomerisation and glycosylation occur in the ER, and the processing of the glycans and palmitoylation in the Golgi (Schmidt, 1982; Gething et al., 1986; Sugrue et al., 1990; Gallagher et al., 1992; Saito et al., 1995). Assembly of the virus components and packaging of the replicated viral genome segments take place at the apical plasma membrane of the polarised host cell (Boulan and Sabatini, 1978; Hughey et al., 1992; Kundu et al., 1996; Lin et al., 1998), preferentially at lipid raft domains from which the virus is released by a budding process (Skibbens et al., 1989; Scheiffele et al., 1999; Zhang et al., 2000). The virion NA prevents the retention of the virus at the target cell membrane (Palese et al., 1974), by cleaving off the Sia from the cell surface. A full set of the eight virus genome segments is required for an infectious virus particle. Two different models have been proposed for the packaging mechanism of the *v*RNA segments. In one of the models, a random incorporation of the *v*RNA segments is suggested, assuming the presence of a common structural domain on all *v*RNPs, and based on experimental evidence that virus particles contain more than eight *v*RNPs, probably forming a significant subpopulation of virions having the full complement of *v*RNA segments (reviewed in Palese and Shaw, 2007). The other model considers the packaging of *v*RNA segments to be a selective process, with each independent *v*RNA segment having a unique

packaging signal and every virion containing a full set of eight *v*RNPs. There is ample experimental support for different features of this model, namely the existence of a packaging signal within coding regions of *v*RNA and specific inter-*v*RNPs interactions (reviewed in Palese and Shaw, 2007).

1.4 The Influenza HA

The HA was originally named because of the ability of the virus to agglutinate erythrocytes (Hirst, 1941). The viral HA mediates the crosslinking of erythrocytes containing the receptor elements identified by the glycoprotein (see below, *1.6*).

1.5.1 HA synthesis and processing

HA is encoded by *v*RNA segment 4 and is synthesised as a single chain precursor polypeptide, HA_0, with a molecular weight (MW) of about 75 kilo-Daltons (kDa), in polyribosomes on the rough ER (Compans, 1973a). The nascent polypeptide is targeted to the ER membrane via an N-terminal signal sequence (McCauley et al., 1979). After cleavage of the signal peptide, the HA_0 corresponds to a type I integral membrane protein, containing a transmembrane stretch near the carboxy C-terminus, the transmembrane anchor, and a short cytoplasmic tail (Figure 3). The insertion of HA_0 in the ER membrane, cleavage of the signal peptide and core glycosylation take place co-translationally (Compans, 1973b). The HA glycosylation was shown to be important for the correct folding and oligomerisation of HA, which forms trimeric molecules (MW = 220 kDa) (Gallagher et al., 1992; Roberts et al., 1993; Hebert et al., 1995; Molinari and Helenius, 2000). Trimerisation is essential for the transition of the HA from the ER to the Golgi network (Gething et al., 1986; Copeland et al., 1988). The cleavage of trimeric HA_0 into the two disulphide-linked chains HA_1 (MW = 50 kDa) and HA_2 (MW = 30 kDa)

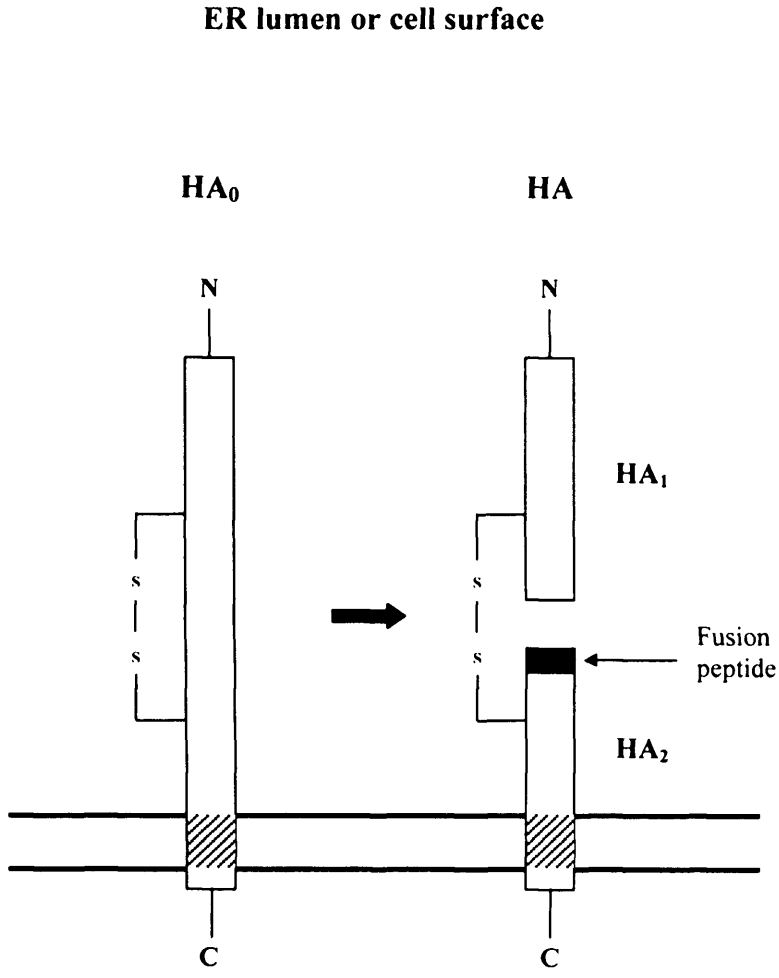


Figure 3. Schematic representation of the orientation and cleavage of the influenza virus integral membrane protein HA

The influenza virus HA is a type I integral membrane protein, having a transmembrane stretch near the carboxy C-terminus, the transmembrane anchor (shown in stripes), and a short cytoplasmic tail. Upon cleavage of the precursor protein HA_0, the fusion peptide is formed at the N-terminus of the HA_2 subunit (see text for details).

(Lazarowitz et al., 1971; Klenk and Rott, 1973) is essential for virus infectivity (Klenk et al., 1975; Lazarowitz and Choppin, 1975). Nonpathogenic viruses are cleaved extracellularly (Lazarowitz et al., 1971; Hay, 1974) by host-specific trypsin-like proteases (Lazarowitz et al., 1973a; Lazarowitz et al., 1973b; Klenk et al., 1977). The enzyme which cleaves influenza viruses in the human respiratory tract is probably similar to the serine protease tryptase Clara, produced by cells of the rat bronchiolar epithelium, which has shown to activate human influenza viruses (Kido et al., 1992). Some highly pathogenic H5 and H7 influenza viruses are cleaved intracellularly (Klenk et al., 1974) by ubiquitous subtilisin-like proteases, such as furin (Stieneke-Grober et al., 1992). The high virulence and systemic infection of these viruses seems to be due to the wide tissue distribution of furin-like enzymes, associated with factors like insertion of polybasic amino acid sequences at the cleavage site or deletion of a carbohydrate in its vicinity, which facilitate the access of the proteases to the cleavage site (reviewed in Steinhauer, 1999). Upon cleavage of HA_0, the newly generated N-terminus of HA_2 forms the fusion peptide, which has a highly conserved hydrophobic amino acid sequence and is implicated in the membrane fusion mechanism (reviewed in Skehel et al., 2000) (see *1.7*).

1.5.2 Crystal structure of the HA ectodomain

The HA ectodomain (HA_1 residues 1-328 and HA_2 residues 1-175) can be released from the viral envelope using the cysteine protease bromelain (Brand and Skehel, 1972), which cleaves off HA_2 residues 176-221, including the transmembrane anchor (Figure 15). This proteolytic isolation of HA from the viral envelope allowed circumventing the aggregation problems associated with the full-length molecule, containing the hydrophobic HA_2 C-terminal membrane-spanning region (Laver and Valentine, 1969). The crystal structure of the bromelain-released HA (BHA) revealed most of the three-dimensional molecular

details of the HA trimer (Wilson et al., 1981). The soluble HA trimer is 135Å long and has two distinct regions: a stem-like core and a globular head at the membrane-distal end of the stem (Figure 4). The stem contains mainly HA_2 residues, which assemble into a 76Å-long triple-stranded coiled coil. The fusion peptide is buried, in an extended conformation, in the interior of the trimer, at 35Å from the membrane-proximal end and 100Å from the membrane-distal. The globular head contains only HA_1 residues forming, in each monomer, an eight-stranded antiparallel β-sheet structure [distorted Jelly roll (Swiss roll) motif], which was later shown to harbour the RBS in the structure of HA with a receptor analogue (Weis et al., 1988) (see below). The fact that the fusion peptide was buried in the native HA trimer pointed to the need of a conformational change, which would lead to its extrusion, rendering the HA molecule fusion active.

The structural information described just above is derived from the X-31 HA, A/Aichi/2/68 strain (see *2.1.1*), the prototype of the H3 viruses responsible for the 1968 HK pandemic (Wilson et al., 1981). More recently, the structure of the cleaved HA ectodomain has been solved for several HAs of different hosts of origin and different subtypes, namely H1, H3, H5, H7 and H9 (Ha et al., 2002; Ha et al., 2003; Gamblin et al., 2004; Russell et al., 2004; Stevens et al., 2004, 2006a; Yamada et al., 2006). These were shown to be structurally very similar [see Table 7 in Appendix I, for a list of root mean square deviation (*r.m.s.d.*) values in the pairwise comparison of the different structures].

1.6 Receptor Binding Mediated by HA

1.6.1 Sialyloligosaccharides as receptors for influenza viruses

A viral infection is initiated by binding of virus particles to specific molecules, the virus receptors, of the target cell surface. The regions of the receptor molecules that directly interact with the virus are defined as receptor determinants (Paulson, 1985). The specific

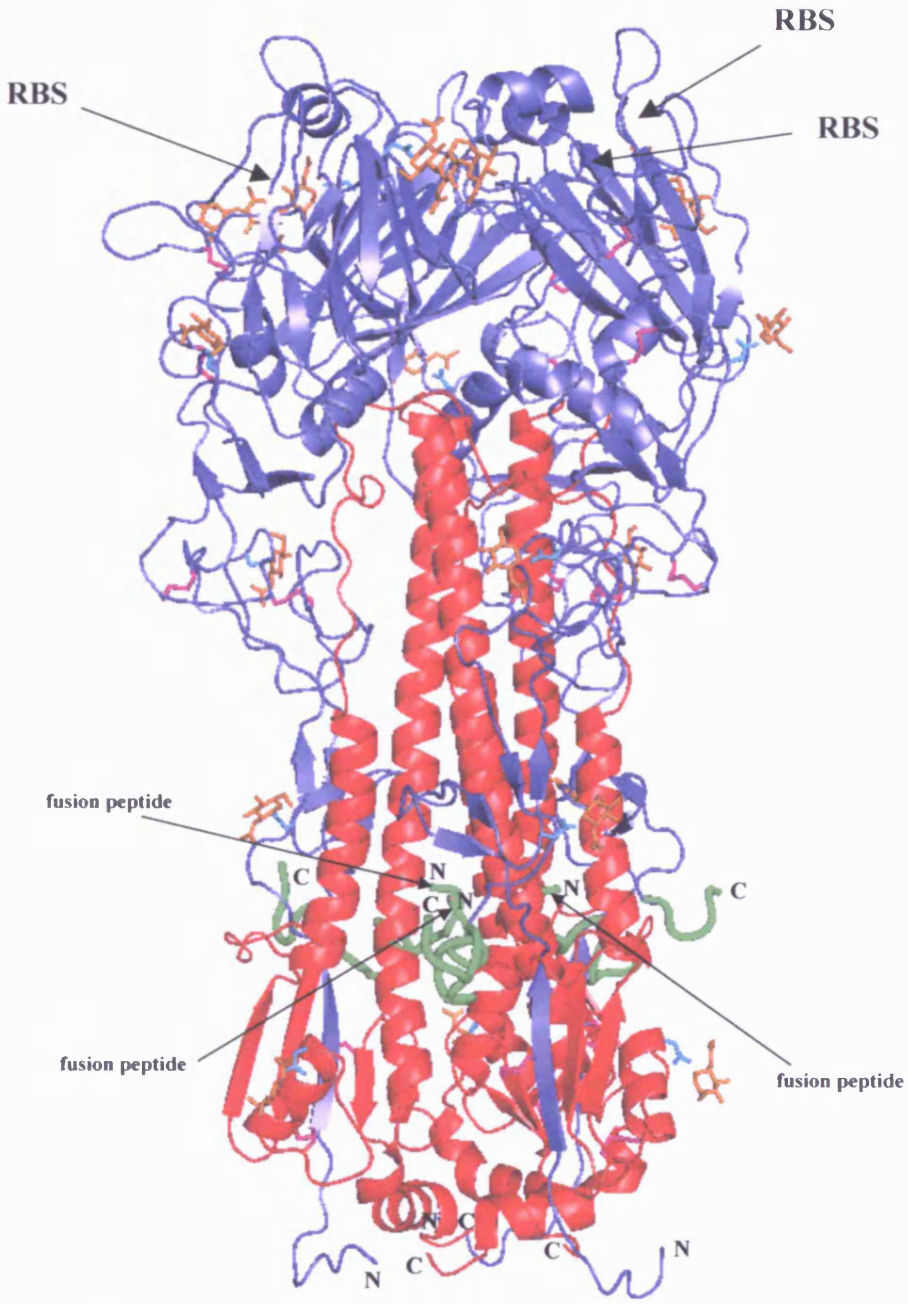


Figure 4. Crystal structure of the native HA ectodomain

Ribbon representation of the crystal structure of the H3 human, X-31 BHA (A/Aichi/2/68). HA_1 is coloured in slate and BHA_2 in red. Residues 323 of HA_1 to 12 of HA_2, corresponding to the HA_0 loop which is cleaved to form the HA molecule primed for activation of the fusion activity (see *1.5.1* and *1.7.1*), are coloured in green and shown with increased width. The locations in the BHA trimer of the fusion peptide and RBSs are indicated. Oligosaccharide molecules and the side chains of the Asn residues at the glycosylation sites are shown in sticks representation and coloured in orange and cyan, respectively. Disulphide bonds are coloured in magenta. The coordinates for X-31 HA were obtained from RCSB/PDB, accession code 1hgf. [The figure was generated using the program Pymol (DeLano, 2006)]

receptors for influenza viruses are still not known, but Sias were a long ago identified as receptor determinants (Klenk et al., 1955; Rosenberg et al., 1956), being directly recognised by the virion HA. Glycoproteins and gangliosides harbour similar sialyloligosaccharide sequences to which influenza viruses bind *in vitro* (reviewed in Paulson, 1985; Suzuki, 1994, 2005), but the relative roles played by glycoproteins and gangliosides in virus infection *in vivo* is not yet determined (discussed by Bukrinskaya, 1982; Paulson, 1985; Herrler et al., 1995).

1.6.1.1 Specificity of influenza viruses for different Sias

Sias are terminal sugar residues on oligosaccharide chains of glycoproteins and glycolipids and are abundant on outer cell membranes and in biological fluids of animals. Due to their chemical variety and wide tissue distribution, they are associated with a multitude of biological functions (Schauer, 1982; Varki, 1992; Schauer et al., 1995; Reuter and Gabius, 1996; Kelm and Schauer, 1997; Schauer and Kamerling, 1997; Varki, 1997).

Chemically, Sias are a family of 9-carbon acid sugars with more than 40 members recognised. Their common distinctive features are a carboxylic group, an amido group and a glycerol tail (C7-C9) attached to a pyranose ring at positions 2, 5 and 6, respectively (Figure 5A). Members of the Sia family differ on the substituents at N5, O4, O7, O8 and O9 (Varki, 1992; 1997; Schauer and Kamerling, 1997).

Two acyl substituents at N5 can be found in nature. The most common derivative is N-acetylneuraminic acid (Neu5Ac), which is a biosynthetic precursor of other Sias (Figure 5B). N-glycolylneuraminic acid (Neu5Gc), derived by enzymatic hydroxylation of the 5-N-acetyl group, is found in all mammals except humans (Schauer, 1982; Varki, 1992). Substitutions of one or more of the hydroxyl groups of Neu5Ac and Neu5Gc with acetyl groups and, less often, methyl, lactyl or sulfate groups, can give rise to other natural Sias.

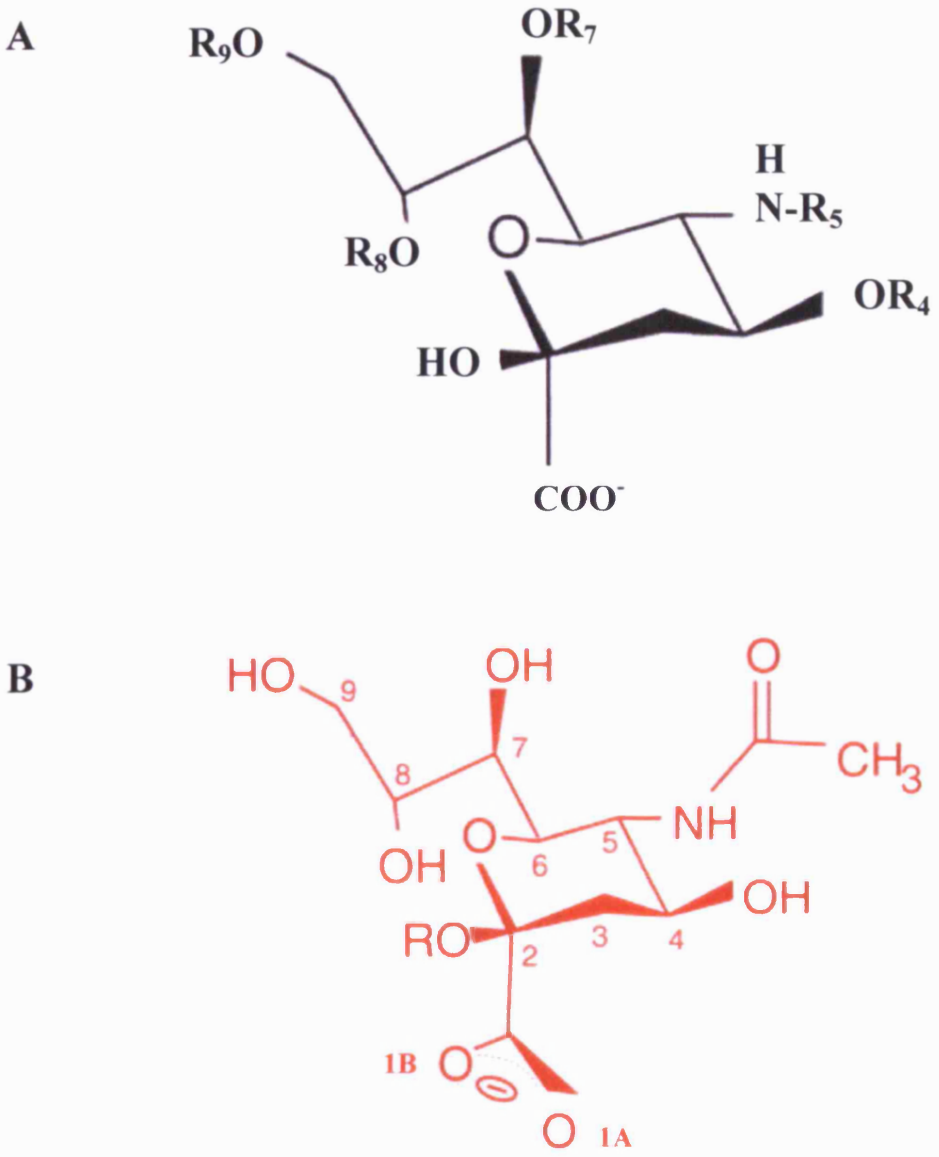


Figure 5. Structure of Sias

(A) Common structural features of Sias from different families
Differences occur at the side-chain groups of the pyranose ring (R)
R_4 = H, acetyl
R_5 = acetyl, glycolyl
R_7 = H, acetyl
R_8 = H, acetyl, methyl, sulfate
R_9 = H, acetyl, L-lactyl, phosphate

(B) Structure of the precursor Sia Neu5Ac (R is the penultimate saccharide residue)

[Adapted from: Skehel and Wiley, 2000]

Of these O-acylated analogues, only the 4-O acetyl, 7-O acetyl and 9-O acetyl-substituted ones were tested for binding to influenza viruses (Levinson et al., 1969; Higa et al., 1985; Pritchett and Paulson, 1989; Sauter et al., 1989; Matrosovich et al., 1992; Sauter et al., 1992a; Matrosovich et al., 1997). The 4-O and 7-O acetylated Sias bind to some influenza virus A strains; 9-O acetylated ones do not bind to influenza A viruses (Higa et al., 1985) but act as receptor determinants for influenza C viruses (Rogers et al., 1986; Herrler and Klenk, 1987; Herrler et al., 1995). The role of other natural Sia derivatives has not been defined. Probably, most of them either cannot support binding of influenza viruses or, due to their presence in low quantities on target cells, do not play a significant role in binding. Therefore, despite the rich chemical diversity of Sias, the non-O-acetylated Neu5Ac and Neu5Gc seem to be the main species used by influenza viruses to infect their natural hosts.

1.6.1.2 Types of α-glycosidic linkage between Sia and penultimate oligosaccharides recognised by influenza viruses

In nature, Sias predominantly exist as components of sialylglycoconjugates and form exclusively α-glycosidic linkages. They are usually attached at the terminal position of oligosaccharide chains and the internal location of Sia residues prevents their binding by influenza viruses (Rogers and Paulson, 1983; Suzuki et al., 1986; Sauter et al., 1989; Suzuki, 1994; Gambaryan et al., 1995; Suzuki et al., 1992). The penultimate sugar residues of oligosaccharide acceptors of Sia are D-galactose (Gal), N-acetyl-D-galactosamine (GalNAc) and, less often, N-acetyl-D-glucosamine (GlcNAc) and Sia. In general, Sias are α2-3- or α2-6-linked to Gal and GalNAc, α2-6-linked to GlcNAc or α2-8-linked to the second Sia residue. Few other types of linkage have been identified, like Sia(α2-4)Gal, Sia(α2-4)GlcNAc and Sia(α2-9)Sia, but these occur less often and their potential recognition by influenza viruses has not been studied yet. In general influenza viruses do

not bind to gangliosides with the terminal compositions of Neu5Ac(α2-8)Neu5Ac or Neu5Ac(α2-8)Neu5Ac(α2-3)Gal(α1-4)Glc in solution (Suzuki et al., 1986, 1992; Matrosovich et al., 1993). Thus, variants of the glycosidic linkage of Sia to the penultimate sugar recognised by influenza viruses in nature are most probably limited to Sia(α2-3/6)Gal, Sia(α2-3/6)GalNAc and Sia(α2-6)GlcNAc. The binding of a H1 virus to α2-8-polysialic acid was recently reported (Wu and Air, 2004), but the potential biological significance of the recognition of α2-8 linkages and the characteristics of this interaction for influenza viruses of different subtypes and hosts of origin have not yet been shown.

1.6.2 Specificity of influenza viruses for Sia-Gal glycosidic linkages

Different binding assays (see *1.6.2.1*) indicated that influenza viruses from different hosts can differentially recognise the terminal Sia(α2-3)Gal and Sia(α2-6)Gal glycosidic linkages of sialyloligosaccharides (see *1.6.8* and *1.6.2.1.1*). In general, human viruses show a preference for Sia(α2-6)Gal-terminated receptors, avian and equine viruses prefer Sia(α2-3)Gal-terminated receptors, and swine viruses bind with similar affinity to both (Rogers and D'Souza, 1989; Connor et al., 1994; Gambaryan et al., 1997; Matrosovich et al., 1997; Ito et al., 1998; reviewed in Wiley and Skehel, 1987 and Skehel and Wiley, 2000). However, several H1 human viruses, as the swine viruses of different subtypes, have a dual binding preference for Sia(α2-3)Gal and Sia(α2-6)Gal glycosidic linkages (Rogers and D'Souza, 1989; Rogers and Paulson, 1983).

1.6.2.1 Characterisation of influenza virus receptor binding specificity and affinity

As the biological receptors for influenza viruses on their target cells have not been

identified, binding assays to characterise the receptor-binding activity of influenza viruses are carried out using more or less well-defined receptor analogues, which are natural and modified Sia-containing molecules or cells (reviewed in Matrosovich et al., 2006).

1.6.2.1.1 Haemagglutination assays using derivatised erythrocytes

Paulson and colleagues introduced a method to generate modified erythrocytes carrying defined sialyloligosaccharide sequences (reviewed in: Paulson, 1985; Paulson and Rogers, 1987). In this procedure, human type B erythrocytes are first treated with *Vibrio cholerae* sialidase to remove Sia, abolishing virus binding. The incubation of the asialo cells with cytidine-5'-monophospho (CMP)-activated Sias and specific sialyltransferases generates modified erythrocytes containing sialyloligosaccharides of defined sequence and Sia content. Hemagglutination assays and virus adsorption studies using the specifically resialylated erythrocytes revealed that different influenza viruses could have a significantly different binding behaviour on the recognition of oligosaccharide sequences and that receptor specificity correlated with species of virus origin (Carroll et al., 1981; Rogers and Paulson, 1983; see *1.6.8*). In these studies, for influenza viruses of the H3 subtype, human isolates with Leu226 had a preference for α2-6-sialyllactosamine (6'SLN) [Neu5Acα2-6Galβ1-4GlcNAc] while avian, equine and some human strains with Gln226 preferred the sequences Neu5Acα2-3Galβ1-3GalNAc, Neu5Acα2-3Galβ1-3GlcNAc and α2-3-sialyllactosamine (3'SLN) [Neu5Acα2-3Galβ1-4GlcNAc]. The sequences Neu5Acα2-3Galβ1-3GalNAc and 6'SLN were determined as being the most appropriate for distinguishing influenza viruses from different hosts, and the sialyltrasferases involved in the generation of those sequences were used in subsequent studies (Higa et al., 1985; Anders et al., 1986; Daniels et al., 1987; Rogers and D'Souza, 1989; Connor et al., 1994; Ito et al., 1998).

1.6.2.1.2 Receptor binding assays using soluble sialosides

The interactions of viruses or individual HA molecules with monovalent sialyloligosaccharides are very weak, with dissociation constants (K_Ds) in the 1-6 mM range (Pritchett et al., 1987; Sauter et al., 1989; Gambaryan et al., 1995, 1999). Therefore, the small differences in the HA binding affinity to sialosides obtained in these assays are significant for receptor specificity (see, below, *1.6.3*).

Different competitive binding assays indicated that the binding affinity of influenza viruses for monovalent siolosides was higher than for free Sia, suggesting the interaction of HA with the penultimate Gal residue (Matrosovich et al., 1993, 1997, 1999, 2000; Gambaryan et al., 1995, 1999). For influenza virus H1 and H3 isolates binding of soluble receptor analogues, avian viruses showed a higher affinity (10 fold) for α2-3-sialyllactose (3'SL) [Neu5Acα2-3Galβ1-4Glc] than for free Sia, indicative of specific interactions of HA with Gal-2; also, avian viruses bind α2-6-sialyllactose (6'SL) [Neu5Acα2-6Galβ1-4Glc] and 6'SLN [Neu5Acα2-6Galβ1-4GlcNAc] with less affinity than to free Sia, suggesting unfavourable interactions with the 6-linked Gal-2 moiety; human and swine viruses, in contrast, bind 3'SL and free Sia with comparable affinity, indicative of a poor interaction with the 3-linked Gal.

In a competitive solid phase assay of influenza virus H1 and H3 isolates binding of sialylglycopolymers and monovalent sialosides, avian viruses showed a high binding affinity for 3'SL [Siaα2-3Galβ1-4Glc], and lower affinity for 6'SLN [Siaα2-6Galβ1-4GlcNAc]; human viruses bound strongly to 6'SLN and did not bind 3'SL; egg-adapted human viruses acquired the ability to bind 3'SL; and the majority of swine viruses had a reduced ability to discriminate between 3'SL and 6'SLN, with a few displaying an avian-like phenotype (Gambaryan et al., 1997).

In an inhibition assay of virus binding to derivatised erythrocytes by soluble sialosides, a

human H3N2 virus showed an eight fold greater affinity for the α2-6-linked LSTc than for the α2-3-linked LSTa (Pritchet et al., 1987) (see *2.2.1*, for a description of the sialyloligosaccharides LSTa and LSTc). In a competitive assay of the binding of another human H3N2 virus strain to soluble sialosides, the virus showed a nine fold higher affinity to LSTc than to LSTa (Gambaryan et al., 1995). In this assay, the binding affinity for 6'SLN was identical to that for LSTc, suggesting that the trisaccharide 6'SLN and the terminal 6'SLN moiety of LSTc have similar conformations (see below, *1.6.7.1*).

In nuclear magnetic resonance (NMR) studies of the binding of X-31 BHA and X-31/HS BHA (L226Q BHA; see *1.6.4*), to 3'SL and 6'SL, the X-31 BHA showed a 1.5 fold higher affinity for 6'SL and the X-31/HS BHA showed a 2.0 fold higher affinity for 3'SL (Sauter et al., 1989; Hanson et al., 1992). In these studies, X-31 BHA showed a slightly lower affinity for LSTa than for 3'SL (Sauter et al., 1989), and the affinity of X-31 virus for 3'SL and 6'SL was similar, but slightly weaker than that for X-31 BHA (Hanson et al, 1992). The similarity between the receptor binding affinities of virus and BHA for simple, soluble sialosides indicates that the RBS is not significantly changed by the bromelain cleavage to release BHA from the virus envelope. Therefore, crystallographic (see *1.6.7*) and NMR studies using BHA accurately reflect the binding of monovalent sialosides to whole virus.

1.6.3 Polyvalency of influenza virus receptor binding

As mentioned above (*1.3.1*), the influenza virion contains multiple HA trimeric spikes (Ruigrok, 1998) and hence is capable of engaging in simultaneous binding of several HA RBSs with multiple copies of the target cell-surface receptors. The affinity of viruses or individual HA molecules for monovalent sialyloligosaccharides is very weak (see, above, *1.6.2.1.2*). The cooperative polyvalent binding of a virus particle allows high binding avidity through multiple low-affinity individual interactions. The measurement of the

binding of HA rosettes (see *3.*) to polyvalent sialosides revealed large increases in binding affinity (Takemoto et al., 1996) and also large differences between the binding of viruses to monovalent and polyvalent sialosides was reported (Gambaryan et al., 1997). The binding of the virus can be increased 10^2-fold by bridging two sialosides with a synthetic spacer (Glick et al., 1991) or greater than 10^3-fold by coupling multiple copies of sialosides to a macromolecular carrier (Matrosovich et al., 1990; Spaltenstine and Whitesides, 1991). Recent advances in the study of whole virus cooperative receptor binding have been obtained using glycan microarrays (reviewed in Stevens et al., 2006b).

1.6.4 Selection of the receptor binding variant X-31/HS, containing the L226Q HA

The HA of the human H3N2 X-31 influenza virus has a binding preference for Neu5Ac(α2-6)Gal-linked sialosides and the variant X-31/HS, which contains the HA with the single amino acid substitution L226Q and prefers Neu5Ac(α2-3)Gal-linked sialosides, was selected by growing X-31 viruses in hen's eggs in the presence of non-immune horse serum (Rogers et al., 1983a). This serum contains the Sia(α2-6)Gal-linkage-rich α_2-macroglobulin, a potent inhibitor of the adsorption and infection of H3 influenza virus isolates (Rogers et al., 1983b; Hanaoka et al., 1989; Pritchett and Paulson, 1989). The change in receptor binding preference induced by the L226Q substitution and its location at the membrane-distal globular head of the HA, confirmed the previous suggestion of a RBS in that region of the molecule, based on the existent homology with the Sia-binding site of wheat-germ agglutinin (Wilson et al., 1981). The high potency inhibition of X-31 virus adsorption and infection by equine α_2-macroglobulin was shown to be due to its high avidity-cooperative binding to the virus (see, above, *1.6.3*) (Hanaoka et al., 1989; Pritchett and Paulson, 1989).

1.6.5 Host range restriction

The contribution of HA for host range restriction was fist shown by Hinshaw et al. (1983) who found that a H3N2 reassortant virus with all six internal genes from an avian strain and the HA or both the HA and NA genes from a human strain could not replicate in the intestinal tract of ducks, rich in α2-3-linked oligosaccharides. Variants of this virus having the mutations L226Q and S228G could efficiently replicate in the duck intestine, however the mutation L226Q was not sufficient for such replication (Naeve et al., 1984; Vines et al., 1998). The L226Q mutation gives the virus the ability to bind α2-3-linked receptors (Rogers et al., 1983a), while the S228G substitution seems to be required for an increased binding affinity (Vines et al., 1998). The importance of the association Leu/Gln and Ser/Gly at residues 226 and 228, respectively, and their correlation with the species of origin in H2 and H3 influenza virus isolates was shown by Connor et al. (1994). All human H3 isolates had Leu226 and Ser228, all equine H3 isolates and practically all avian isolates had Gln226 and Gly228. This correlation extended to the H2 isolates. Also, all human isolates were very sensitive in haemagglutination assays using native and resialylated erythrocytes to inhibition by the α2-6-linkage-rich α_2-macroglobulin of horse sera (see, above, *1.6.4*), while avian and equine viruses were resistant. H3 equine viruses show stricter receptor specificity for α2-3-linked receptors than duck viruses. Some duck viruses are double-binders, also recognising α2-6-linked receptors (Rogers and Pauson, 1983; Connor et al., 1994). The two properties, linkage specificity and sensitivity or resistance to horse serum, together define the receptor specificity of H3 virus isolates (Connor et al., 1994).

The L226Q mutation in the HA of H3 and H2 viruses is essential for the ability of human viruses to initially infect the intestinal tract of ducks. Irrespective of the amino acid at position 228, HAs having Leu at position 226 show no evidence of avian receptor binding

characteristics (Kobasa et al., 1997). The avian virus receptor-binding phenotype requires the presence of Gln at position 226 (Vines et al., 1998). The selection of the receptor-binding variant of the H3 avian virus A/duck/Ukraine/1/63, having the Q226L mutation at the RBS, the reverse of the L226Q mutation of X-31/HS, yielded a virus with a shifted preference for Sia(α2-6)Gal linkages (Rogers et al., 1985). But this variant reverted the receptor-binding phenotype during a single passage in embryonated chicken eggs. As the virus was grown in the allantoic cavity, which is devoid of Sia(α2-6)Gal linkages, this suggests that in contrast to X-31 and human H3 virus isolates, the H3 avian Q226L mutant is not able to bind Sia(α2-3)Gal linkages efficiently.

Therefore, position 226 of influenza H2 and H3 HA seems to play a pivotal role in the transmission of viruses between different hosts, by defining the linkage specificity of the virus. Other factors, like the additional mutation at position 228 (Vines et al., 1998), and additional viral genes will then contribute for the completion of the interspecies transmission process (Baigent and McCauley, 2003).

1.6.6 Structure of the HA RBS

The receptor-binding pocket that accommodates the Sia residue is a shallow depression at the surface of the membrane-distal globular head in each monomer of the HA trimer (Weis et al., 1988). A group of amino acid residues that are conserved among different virus strains form the RBS, and a set of amino acid residues that change in response to the host immune response (antigenic drift; see below, *1.6.9*) surround the site (Skehel and Wiley, 2000). Among the conserved residues in the RBS region, some may be essential for direct interactions with the receptor, others may be important for the structural integrity of the HA molecule and/or that of the RBS.

For the X-31 HA RBS, at the bottom of the pocket, residues Tyr98, Trp153, His183 and Tyr195 form a hydrogen bond network (Figure 6). Residues 187-194 form a short α-helix, the 190 helix, and the side chains of residues Glu190 and Leu194 together with those of His183 and Thr155 form the rear edge of the site. Residues 134-138 form a surface loop, the 130 loop, which defines the right side of the RBS. The left edge of the Sia-binding pocket is defined by the stretch of residues 224-228, the 220 loop.

The Leu226 to Gln226 mutation in the L226Q HA was found to induce a slight narrowing in the structure of the RBS in comparison with that of the X-31 HA (Weis et al., 1988; 1990). The loss of a van der Waals contact between Leu226 and Ala138 and the formation of a new hydrogen bond from Gln226 to Ser136 seem to form the basis for the new RBS geometry in the L226Q HA mutant, with the consequent upward projection of Gln226 side chain. Smaller shifts in residue position were transmitted to regions in and around the L226Q HA RBS (Weis et al., 1988).

The structural differences of the RBS of H3 avian HA, A/duck/Ukraine/63, in comparison with that of the X-31 HA are similar to the differences observed between the RBSs of X-31 HA and L226Q HA. Of the twenty amino acid differences between X-31 HA and H3 avian HA, only ten are in the membrane-distal globular head domain, and of those only five, N137S, S193N, L226Q, S227P, S228G, are at the RBS. The L226Q mutation seems to induce in the X-31 HA most of the positional differences seen when comparing the H3 avian HA RBS with the X-31 HA RBS. The X-31 HA has Leu226 and both L226Q HA and H3 avian HA have Gln226. However, residues 227 and 228 apparently are also determinants for the conformation of the RBS. The X-31 HA and L226Q HA have Ser227, but the H3 avian HA has the unusual Pro227. Both the X-31 HA and L226Q HA have Ser228, but the H3 avian HA has the typically avian Gly228. The residual difference between the conformations of the RBSs of L226Q HA and H3 avian HA are probably due

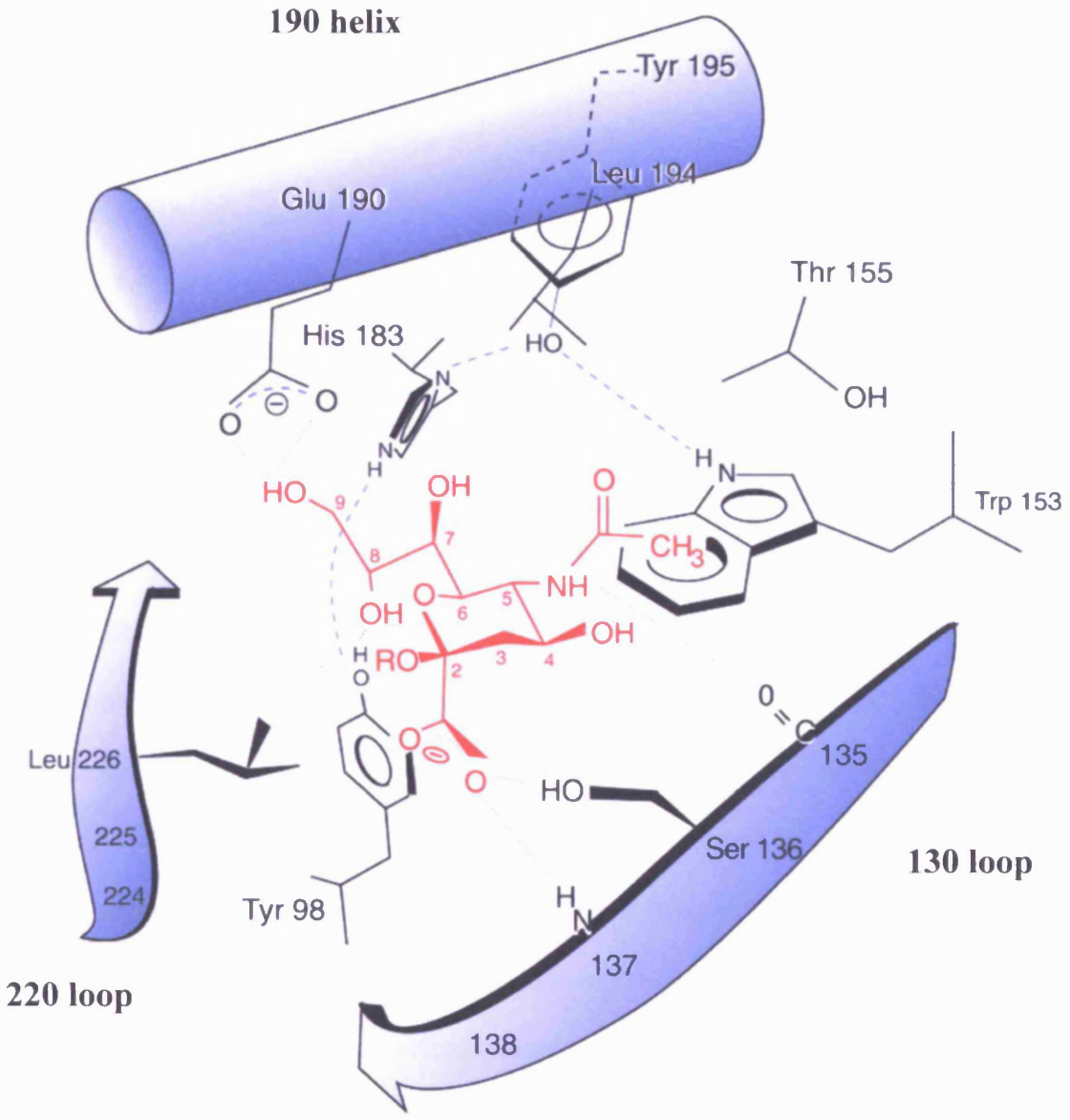


Figure 6. The HA RBS

The HA RBS in complex with Neu5Ac. The three main secondary structural elements of the HA RBS - 130 loop, 190 helix and 220 loop - are indicated in blue. Potential hydrogen bonds between conserved residues at the bottom of the site are indicated by dashed lines. The terminal Neu5Ac of a receptor analogue bound on the RBS is shown in red. Putative hydrogen bonds between the HA and the Neu5Ac moiety are shown as dotted lines. The RBS shown is that of the X-31 HA. The HA RBS forms a generally conserved set of interactions with the Sia moiety, irrespective of the receptor analogue. [Adapted from: Skehel and Wiley, 2000]

to the H3 avian P227 and G228 residues, which seem to allow the 220 loop to be closer to Y98 than the residues S227 and S228 of the L226Q mutant (the distance of the phenol OH group of Tyr98 to the Cα of residue 228 is 3.7 Å in the H3 avian HA, but 4.6 Å in the L226Q HA). The superposition of the 220 loop in the structures of L226Q HA, X-31 HA and H3 avian HA, revealed that it has a *r.m.s.d.* on all C_α positions of 0.13 Å, 0.31 Å and 0.36 Å in the pairwise comparisons L226Q HA/X-31 HA, L226Q HA/H3 avian HA and H3 avian HA/X-31 HA, respectively (these values were calculated as indicated in *2.2.4.4.1*). The A/duck/Ukraine/63 virus is considered to be a possible progenitor of the virus that caused the 1968 HK human flu pandemic (Ha et al., 2003), namely due to the amino acid sequence and structural similarity of the H3 avian HA to the H3 human X-31 HA (see, also, *1.6.7.1*).

The conformations of each of the conserved secondary structural elements of the HA RBS (above-described; Figure 6) are similar in all HAs studied thus far, with small differences accounting for variations in the mode of binding to different receptor analogues (see *1.6.7.1*). The amino acids present at HA RBS positions 226 and 228, in the 220 loop, in H2 and H3 HAs have been shown to be correlated with the shift in viral preference between avian and human receptors and therefore related to viral interspecies transmission (Connor et al., 1994; Naeve et al., 1984). In X-31 HA, they are Leu226 and Ser228; in L226Q HA, Gln226 and Ser228; in H3, H5, and H7 avian HAs, and H1 swine, H1 human and 1918 HAs, they are Gln226 and Gly228; in H9 swine HA, Leu226 and Gly228. The H1 HAs have the particularity of efficiently binding human receptor analogues even though they have a Gln at position 226 (Gamblin et al, 2004). They have a particular RBS conformation, with the side chain of Gln226 positioned lower in the site by comparison with HAs of other subtypes also having Gln at the same position. The L226Q HA of the present study and the H9 swine HA (Ha et al., 2001) seem to have an intermediate RBS

between avian and human. When compared with the above-mentioned HAs, except the H1 HAs, the L226Q HA has an avian Gln226 but a human Ser228, and the H9 swine HA has a human Leu226 and an avian Gly228. Even though positions 226 and 228 of the H2 and H3 HA RBS seem to play a major role in the virus receptor specificity, as observed in studies of virus-mediated haemagglutination and virus replication in the colon cells of ducks (Rogers and Paulson, 1983; Connor et al., 1994; Vines et al., 1998), there are RBS mutations at positions other than 226 and 228 that can affect the viral receptor binding specificity, as described for the replication of influenza viruses in cells or tissues from different species (Gubareva et al., 1994; Hardy et al., 1995; Robertson et al., 1995). Crystal structures of HA-receptor-analogue complexes that could reveal the molecular basis at the RBS for those changes in receptor recognition are not yet described.

Two recent studies using a solid phase binding assay and glycan microarrays also revealed the importance of several different mutations in different regions of the H5 avian HA molecule for the viral receptor specificity (Stevens et al., 2006a; Yamada et al., 2006).

1.6.7 Crystal structures of HA in complex with receptor analogues

The crystal structures of BHA in complex with 3'SL, 6'SL, and different synthetic Sia derivatives have been determined by X-ray crystallography (Weis et al., 1988, 1990; Sauter et al., 1992a, 1992b; Watowich et al., 1994; Eisen et al., 1997). Most of the studies were done using the X-31 BHA (reviewed in Wiley and Skehel, 1987; Skehel and Wiley, 2000). More recently, these studies have involved the structures of HAs of different subtypes and hosts of origin in complex with the receptor analogues LSTa and LSTc (see below, *1.6.7.1*), having been extended to the H3, H5 and H7 avian HAs, H1 and H9 swine HAs, and H1 human HA (Eisen et al., 1997; Ha et al., 2001, 2003; Gamblin et al., 2004; Russell et al., 2006). All these structures have a biological significance, at the level of

receptor binding, in terms of transmission of influenza viruses to the human population. The H3 and H1 HAs are connected with pandemics (see below, *1.6.9*) (Eisen et al., 1997; Gamblin et al., 2004), H5 avian HA and H9 swine HA with outbreaks in HK in 1997 and 1999 (Ha et al., 2001), and the H7 avian HA with outbreaks in the Netherlands in 2003 and more recently in North America (Russell et al., 2006).

The structure and orientation of Sia bound on the X-31 HA RBS in all hitherto studied HA–receptor-analogue complexes is very similar, regardless of the receptor analogue (Figure 6). One side of the pyranose ring forms an interface with the base of the site and atoms of the ring substituents, namely the axial carboxylate, the acetamido nitrogen and the 8- and 9- hydroxyl groups of the glycerol arm, face main-chain and side-chain polar atoms of identically-positioned residues to form hydrogen bonds. More precisely, the side chain of conserved Ser136 and the amide of peptide bond 137 are hydrogen bonded to the carboxylate; the carbonyl of peptide bond 135 forms a hydrogen bond with the 5-acetamido nitrogen; the side chains of His183 and Glu190 hydrogen bond to the 9-hydroxyl; Tyr98 forms a hydrogen bond with 8-hydroxyl. The acetamido methyl group forms a van der Waals contact with the six-membered ring of Trp153, and the 7-hydroxyl and acetamido carbonyl bond to each other and form a van der Waals contact with Leu194. The 4-hydroxyl group has not shown any contribution for binding. Results from binding studies and site-directed mutations correlate well with these structural data and confirmed the importance of different residues for the interaction with receptor analogues (reviewed in Skehel and Wiley, 2000). The just-described interactions between X-31 HA and the Sia of different sialosides were shown as being generally conserved in the crystal structures of HAs of different subtypes in complex with LSTa and LSTc (see, below, *1.6.7.1*). The contacts between individual HAs and the asialo part of the receptor analogue are important for the definition of recognition patterns related with the virus receptor binding specificity.

1.6.7.1 HA in complex with the receptor analogues LSTa and LSTc

Sialyloligosaccharides bearing terminal Sia(α2-3)Gal or Sia(α2-6)Gal linkages have different configurations around the glycosidic bond that can be differentially recognised by the influenza HA. In the crystal structures of LSTa and LSTc (see *2.2.1*, for a description of the sialyloligosaccharides LSTa and LSTc) with HAs of different hosts and subtypes (Eisen et al., 1997; Ha et al., 2001, 2003; Gamblin et al., 2004; Russell et al., 2006) the Sia(α2-3)Gal linkage of LSTa is found in the configurations *cis* or *trans* and the Sia(α2-6)Gal linkage of LSTc in the configuration *cis* (Figures 7-13), which correlate with the configurations *syn* (*cis*) and *anti* (*trans*) determined for α2-3-linked and α2-6-linked sialosides in solution by modelling, conformational energy calculations and NMR studies (Breg et al., 1989; Poppe et al., 1989, 1992; Sabesan et al., 1991). Figures 7-13 were generated using the program Pymol (DeLano, 2006).

The binding conformation of the avian receptor analogue LSTa is very similar in H3 avian HA, H5 avian HA, H7 avian HA and H1 human HA, having the *trans* (or *anti*) configuration about the α2-3-glycosidic linkage (Figures 8A, 9A, 10 and 12A). H3 human HA and H9 swine HA bind LSTa in the *cis* configuration, and H1 swine HA seems not to bind the avian receptor efficiently (Figures 7A, 11A, and 13A). The H3 human HA, H9 swine HA, H1 human HA and H1 swine HA bind LSTc in the *cis* (or *syn*) configuration about the glycosidic bond, the H5 avian HA binds it poorly and the H7 avian HA does not bind LSTc. However, despite the clear preference of H3 avian HA for LSTa, it can still bind LSTc in the *cis* configuration, as observed for the H3 human HA (Figures 7B and 8B). This receptor linkage recognition by the H3 avian HA gives support to the proposal of A/duck/Ukraine/63 virus as the possible progenitor of the virus that caused the 1968 HK human flu pandemic (Ha et al., 2003). A comparison of the linkage type and configuration of the Sia-Gal glycosidic linkage preferred by the HAs of different subtypes and hosts of

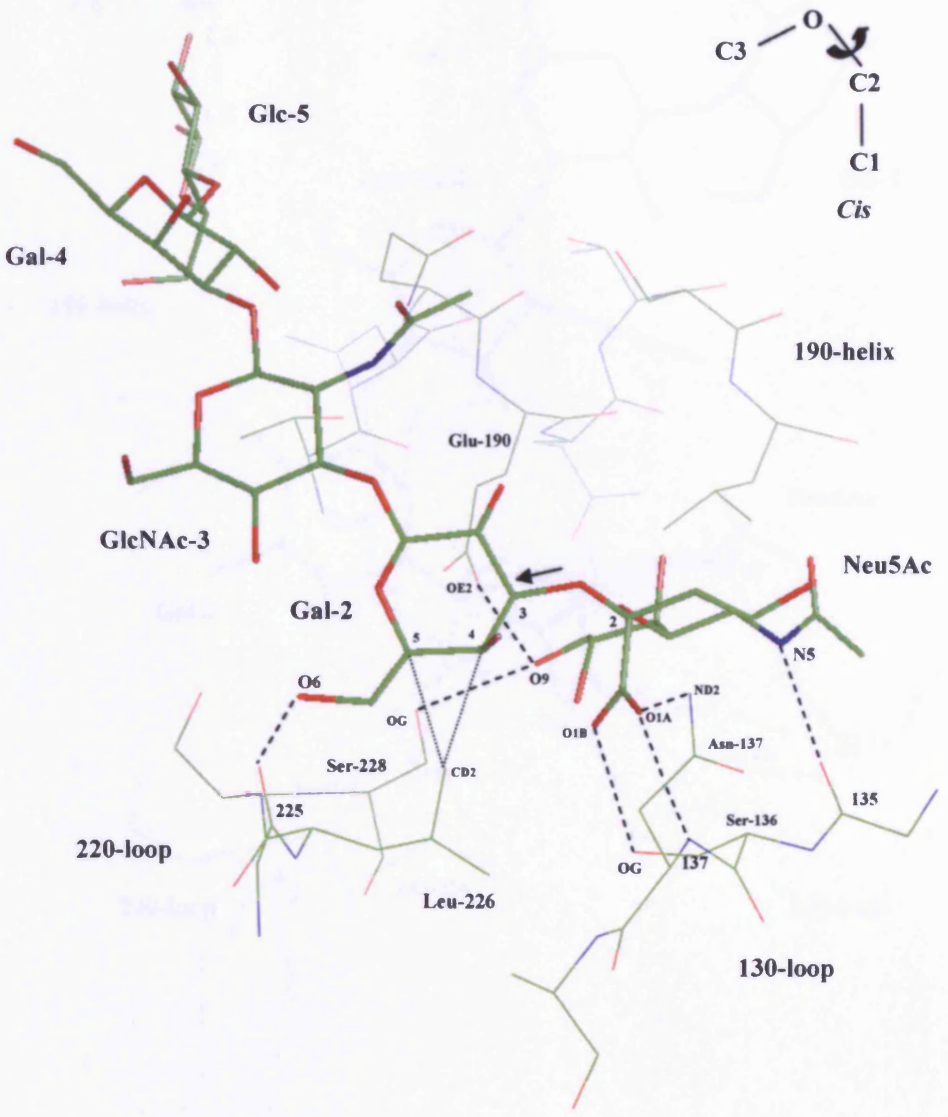


Figure 7A. Structure of H3 human HA (X-31) in complex with LSTa

The conformation of LSTa bound on the RBS of X-31 HA is shown. The electron density allowed the location of all five saccharide residues of LSTa. Selected potential hydrogen bonds between analogue and X-31 HA are shown as dashed lines. Atoms are coloured as: green for carbon, blue for nitrogen, and red for oxygen. Relevant residues of the HA RBS and atoms of LSTa and the HA RBS, namely those involved in the selected interatomic contacts, are specifically identified (see Table 10, Appendix IV). The three main secondary structural elements of the HA RBS - 130 loop, 190 helix and 220 loop - are also indicated. The arrow indicates the configuration around the glycosidic bond, as shown schematically at the top right of the figure. The side chain of Leu226 makes van der Waals contacts with the Gal-2 of LSTa (dotted lines) (Eisen et al., 1997). [The coordinates for X-31 HA-LSTa were provided by M. B. Eisen]

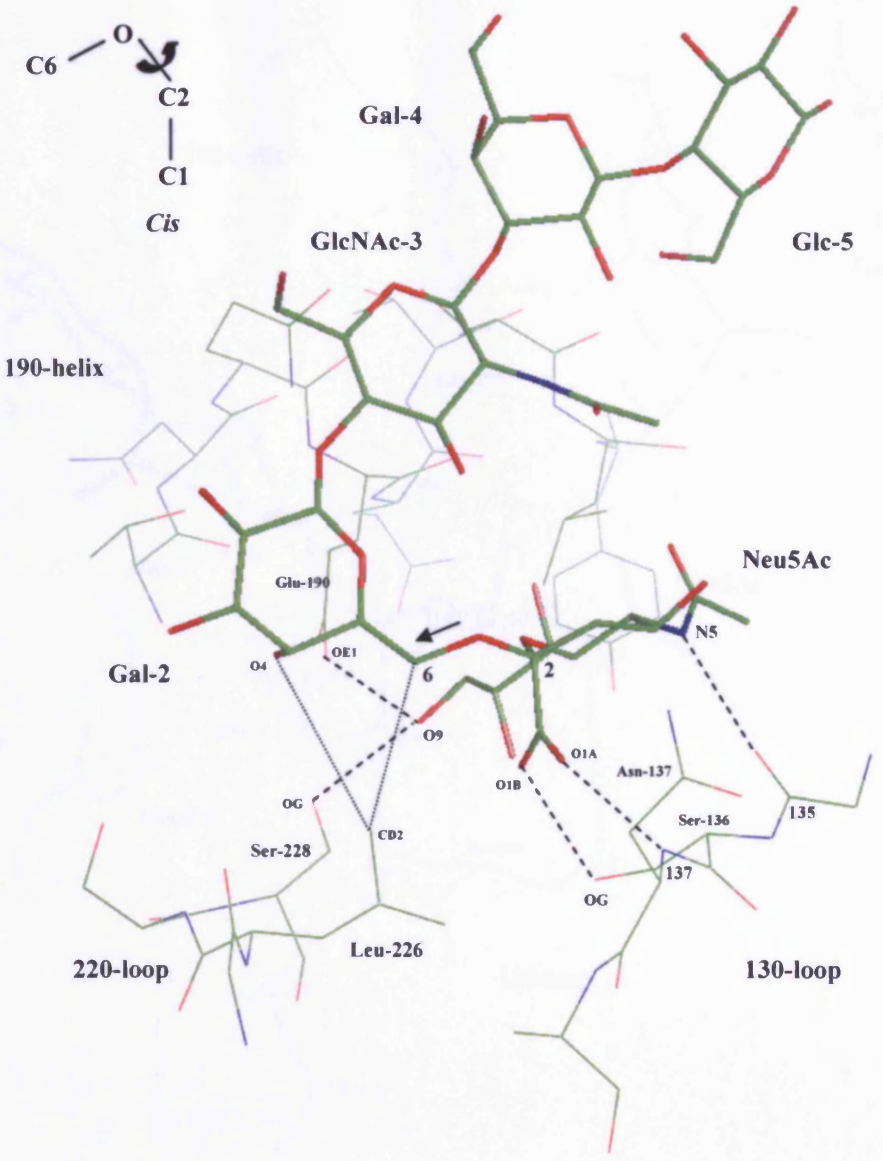


Figure 7B. Structure of H3 human HA (X-31) in complex with LSTc

The conformation of LSTc bound on the RBS of X-31 HA is shown. The electron density allowed the location of all five saccharide residues of LSTc. Selected potential hydrogen bonds between analogue and X-31 HA are shown as dashed lines. Atoms are coloured as: green for carbon, blue for nitrogen, and red for oxygen. Relevant residues of the HA RBS and atoms of LSTc and the HA RBS, namely those involved in the selected interatomic contacts, are specifically identified (see Table 11, Appendix IV). The three main secondary structural elements of the HA RBS - 130 loop, 190 helix and 220 loop - are also indicated. The arrow indicates the configuration around the glycosidic bond, as shown schematically at the top left of the figure. The side chain of Leu226 makes van der Waals contacts with the Gal-2 of LSTc (dotted lines) (Eisen et al., 1997). [The coordinates for X-31 HA-LSTc were provided by M. B. Eisen]

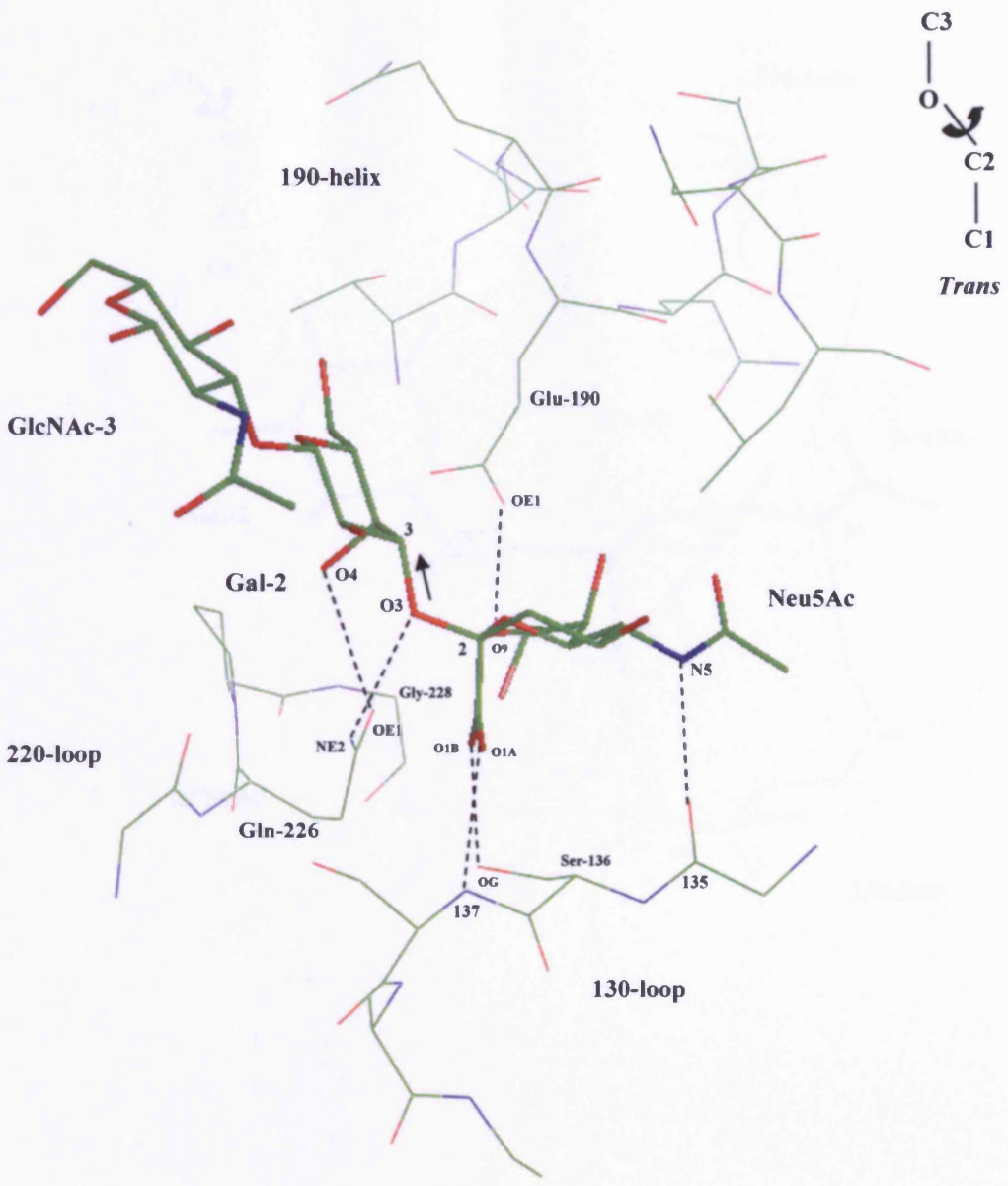


Figure 8A. Structure of H3 avian HA (Dk/Ukraine/1/63) in complex with LSTa

The conformation of LSTa bound on the RBS of H3 avian HA is shown. The electron density allowed the location of only the first three saccharide residues of LSTa. Selected potential hydrogen bonds between analogue and H3 avian HA are shown as dashed lines. Atoms are coloured as: green for carbon, blue for nitrogen, and red for oxygen. Relevant residues of the HA RBS and atoms of LSTa and the HA RBS, namely those involved in the selected interatomic contacts, are specifically identified (see Table 10, Appendix IV). The three main secondary structural elements of the HA RBS - 130 loop, 190 helix and 220 loop - are also indicated. The arrow indicates the configuration around the glycosidic bond, as shown schematically at the top right of the figure (Ha et al., 2003). [The coordinates for H3 avian HA-LSTa were obtained from the Research Collaboratory for Structural Bioinformatics (RCSB)/Protein Data Bank (PDB), accession code 1mqm]

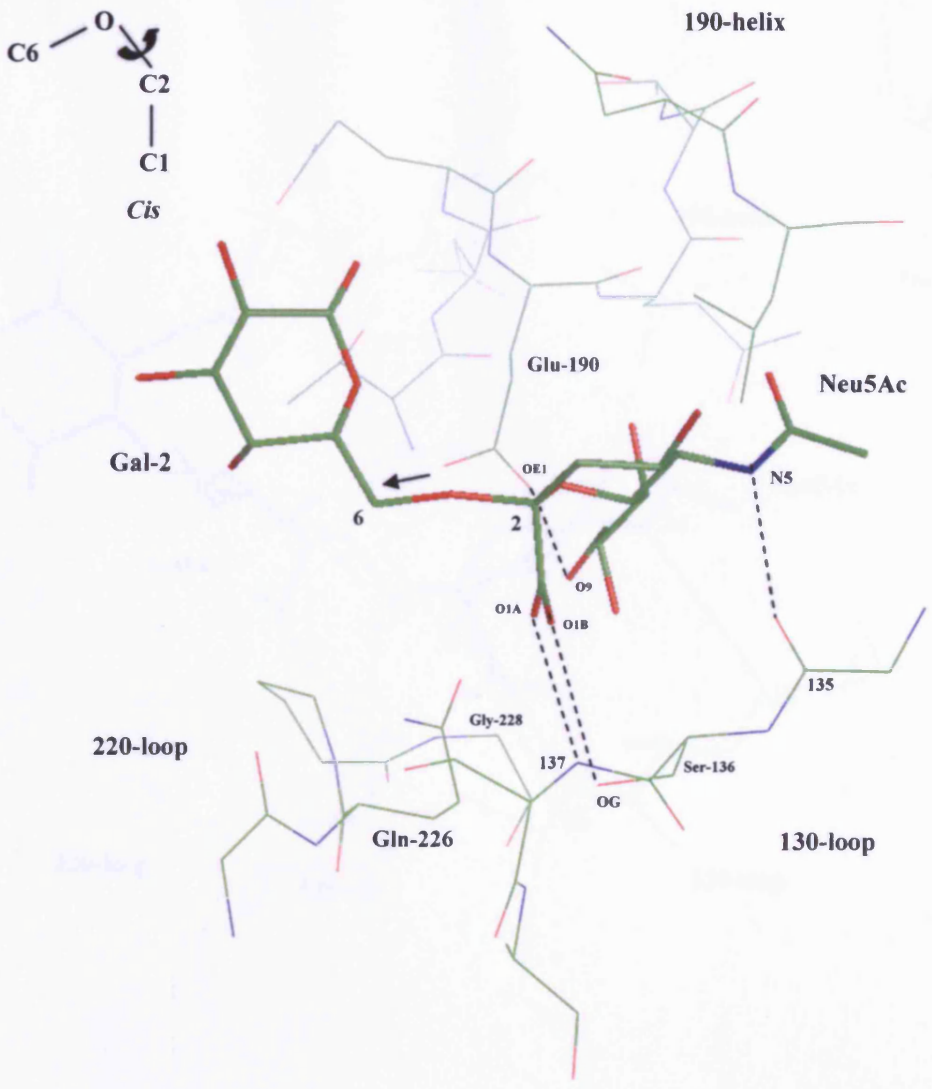


Figure 8B. Structure of H3 avian HA (Dk/Ukraine/1/63) in complex with LSTc

The conformation of LSTc bound on the RBS of H3 avian HA is shown. The electron density allowed the location of only the first two saccharide residues of LSTc. Selected potential hydrogen bonds between analogue and H3 avian HA are shown as dashed lines. Atoms are coloured as: green for carbon, blue for nitrogen, and red for oxygen. Relevant residues of the HA RBS and atoms of LSTc and the HA RBS, namely those involved in the selected interatomic contacts, are specifically identified (see Table 11, Appendix IV). The three main secondary structural elements of the HA RBS - 130 loop, 190 helix and 220 loop - are also indicated. The arrow indicates the configuration around the glycosidic bond, as shown schematically at the top left of the figure (Ha et al., 2003). [The coordinates for H3 avian HA-LSTc were obtained from RCSB/PDB, accession code 1mqn]

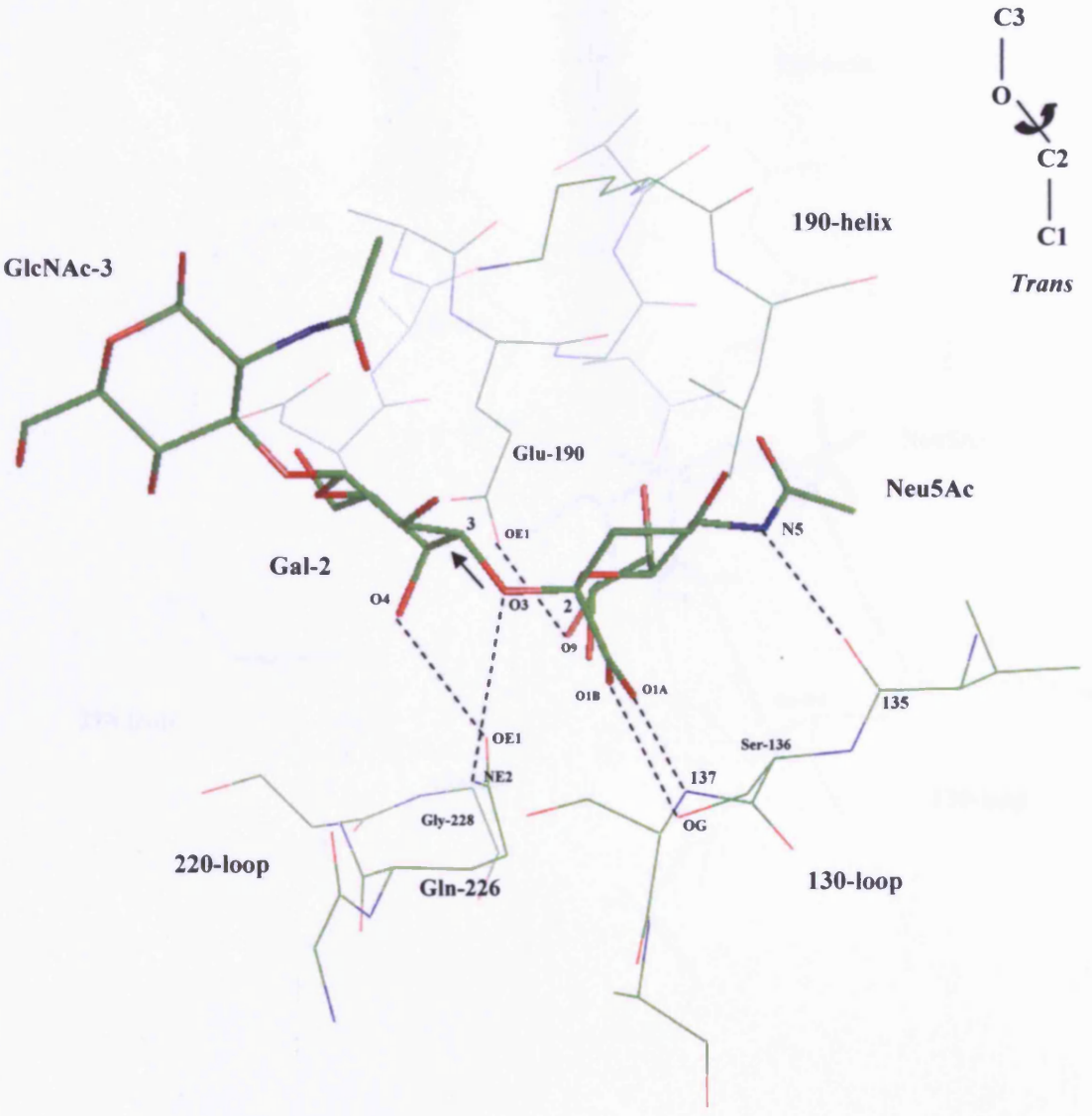


Figure 9A. Structure of H5 avian HA (Dk/Singapore/3/97) in complex with LSTa

The conformation of LSTa bound on the RBS of H5 avian HA is shown. The electron density allowed the location of only the first three saccharide residues of LSTa. Selected potential hydrogen bonds between analogue and H5 avian HA are shown as dashed lines. Atoms are coloured as: green for carbon, blue for nitrogen, and red for oxygen. Relevant residues of the HA RBS and atoms of LSTa and the HA RBS, namely those involved in the selected interatomic contacts, are specifically identified (see Table 10, Appendix IV). The three main secondary structural elements of the HA RBS - 130 loop, 190 helix and 220 loop - are also indicated. The arrow indicates the configuration around the glycosidic bond, as shown schematically at the top right of the figure (Ha et al., 2001). [The coordinates for H5 avian HA-LSTa were obtained from RCSB/PDB, accession code 1jsn]

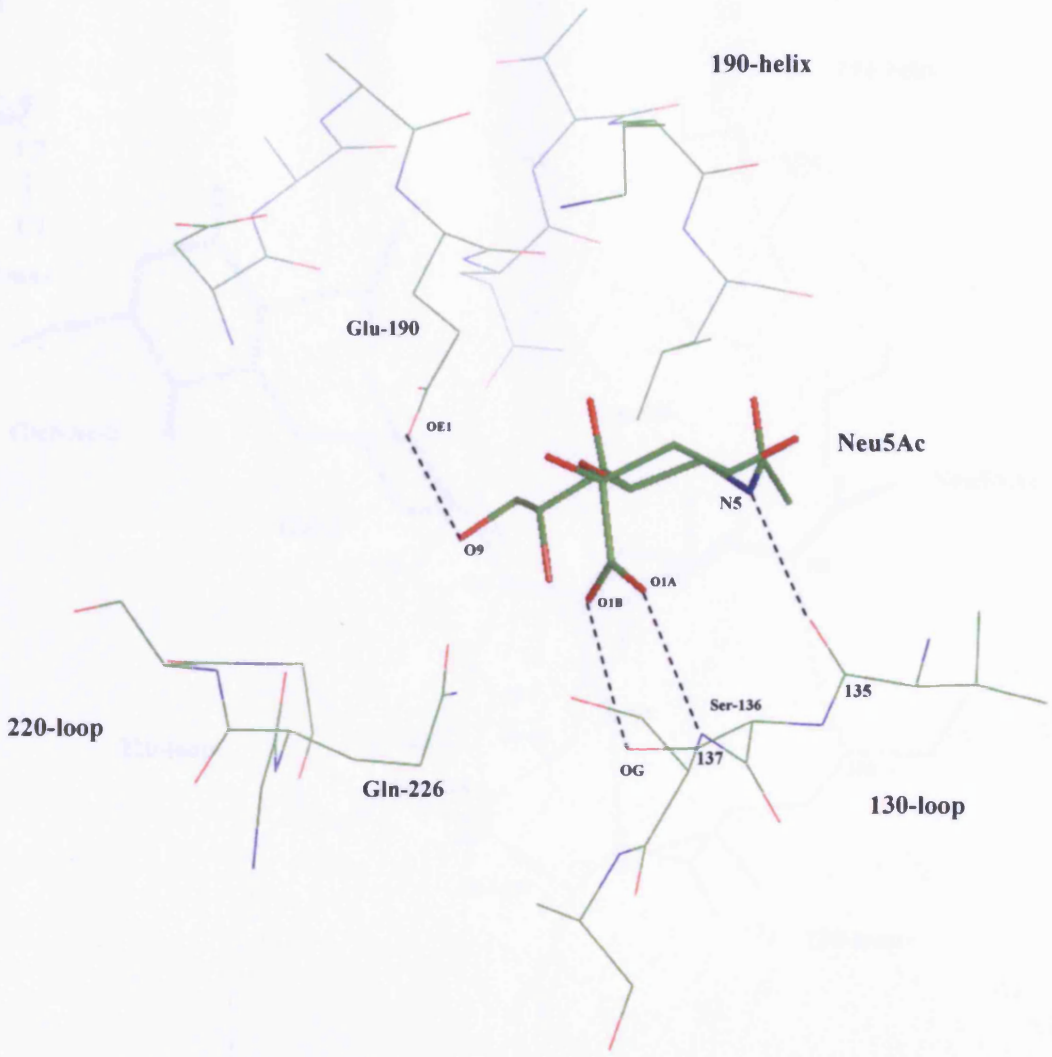


Figure 9B. Structure of H5 avian HA (Dk/Singapore/3/97) in complex with LSTc

Only the Sia of LSTc, bound on the RBS of H5 avian HA, could be fitted in the electron density. Selected potential hydrogen bonds between analogue and H5 avian HA are shown as dashed lines. Atoms are coloured as: green for carbon, blue for nitrogen, and red for oxygen. Relevant residues of the HA RBS and atoms of Sia and the HA RBS, namely those involved in the selected interatomic contacts, are specifically identified (see Table 11, Appendix IV). The three main secondary structural elements of the HA RBS - 130 loop, 190 helix and 220 loop - are also indicated (Ha et al., 2001). [The coordinates for H5 avian HA-LSTc were obtained from RCSB/PDB, accession code 1jso]

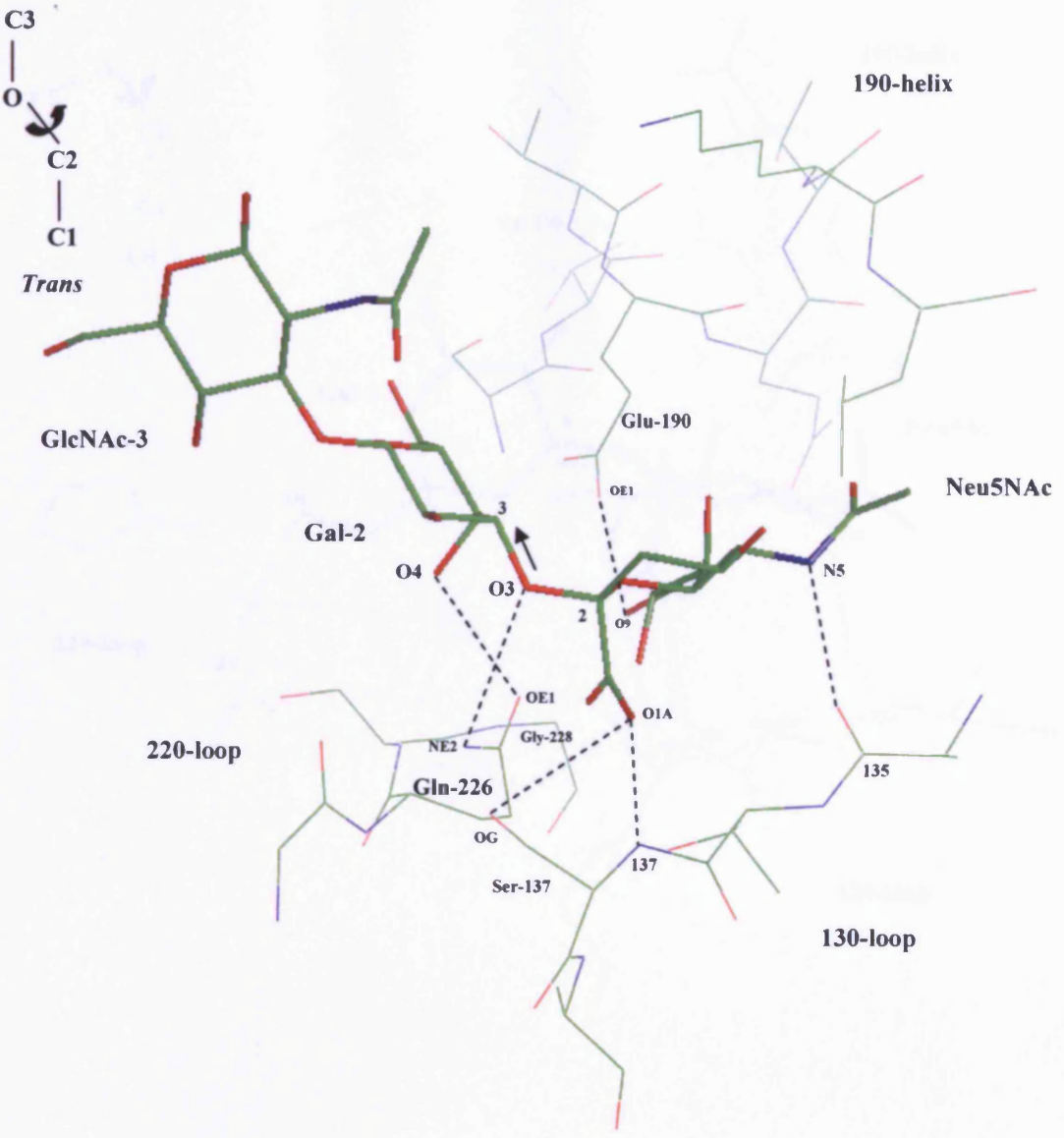


Figure 10. Structure of H7 avian HA (Turkey/Italy/02) in complex with LSTa

The conformation of LSTa bound on the RBS of H7 avian HA is shown. The electron density allowed the location of only the first three saccharide residues of LSTa. Selected potential hydrogen bonds between analogue and H7 avian HA are shown as dashed lines. Atoms are coloured as: green for carbon, blue for nitrogen, and red for oxygen. Relevant residues of the HA RBS and atoms of LSTa and the HA RBS, namely those involved in the selected interatomic contacts, are specifically identified (see Table 10, Appendix IV). The three main secondary structural elements of the HA RBS - 130 loop, 190 helix and 220 loop - are also indicated. The arrow indicates the configuration around the glycosidic bond, as show schematically at the top left of the figure (Russell et al., 2006). [The coordinates for H7 avian HA-LSTa were provided by R. J. Russell]

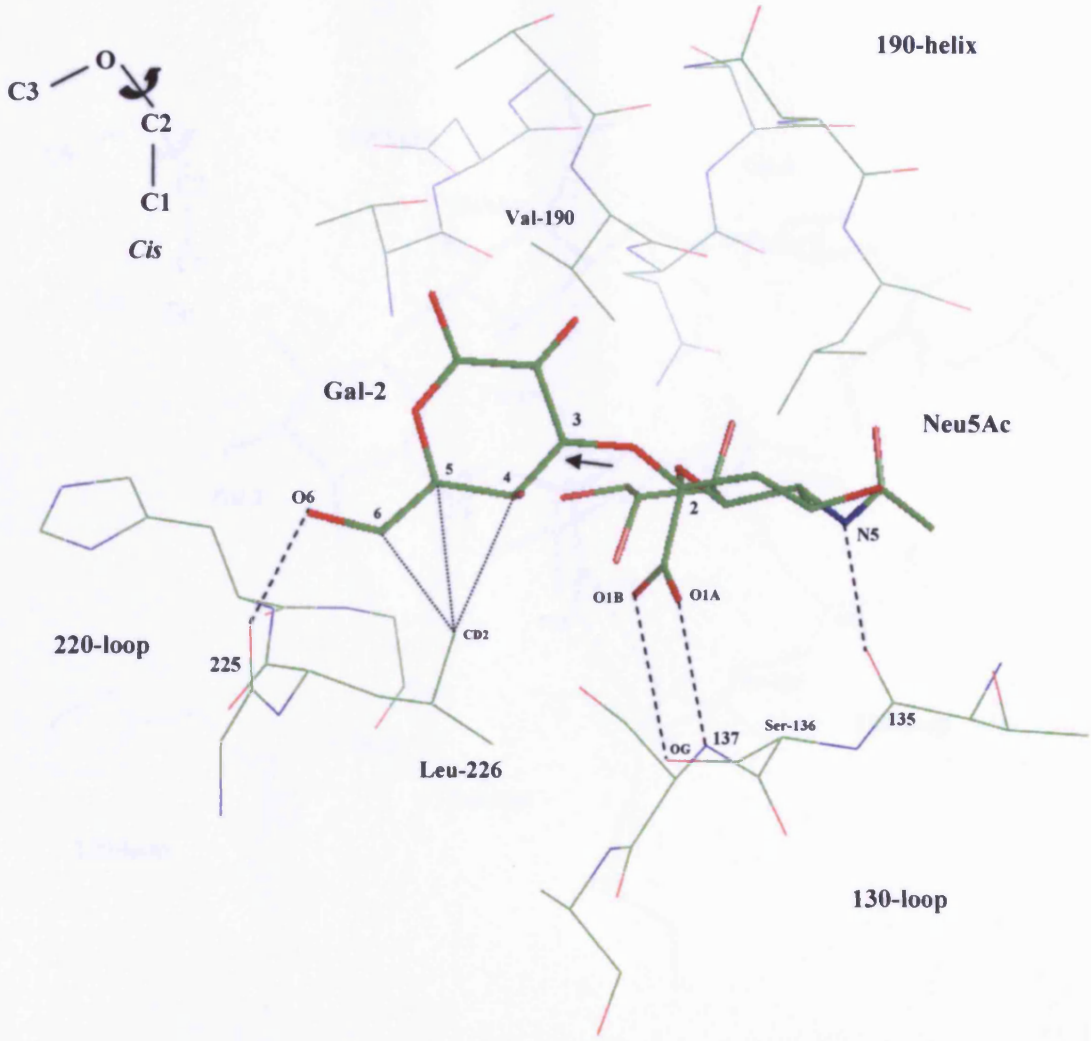


Figure 11A. Structure of H9 swine HA (Sw/Hong Kong/9/98) in complex with LSTa

The conformation of LSTa bound on the RBS of H9 swine HA is shown. The electron density allowed the location of only the first two saccharide residues of LSTa. Selected potential hydrogen bonds between analogue and H9 swine HA are shown as dashed lines. Atoms are coloured as: green for carbon, blue for nitrogen, and red for oxygen. Relevant residues of the HA RBS and atoms of LSTa and the HA RBS, namely those involved in the selected interatomic contacts, are specifically identified (see Table 10, Appendix IV). The three main secondary structural elements of the HA RBS - 130 loop, 190 helix and 220 loop - are also indicated. The arrow indicates the configuration around the glycosidic bond, as shown schematically at the top left of the figure. The side chain of Leu226 makes van der Waals contacts with the Gal-2 of LSTa (dotted lines) (Ha et al., 2001). [The coordinates for H9 swine HA-LSTa were obtained from RCSB/PDB, accession code 1jsh]

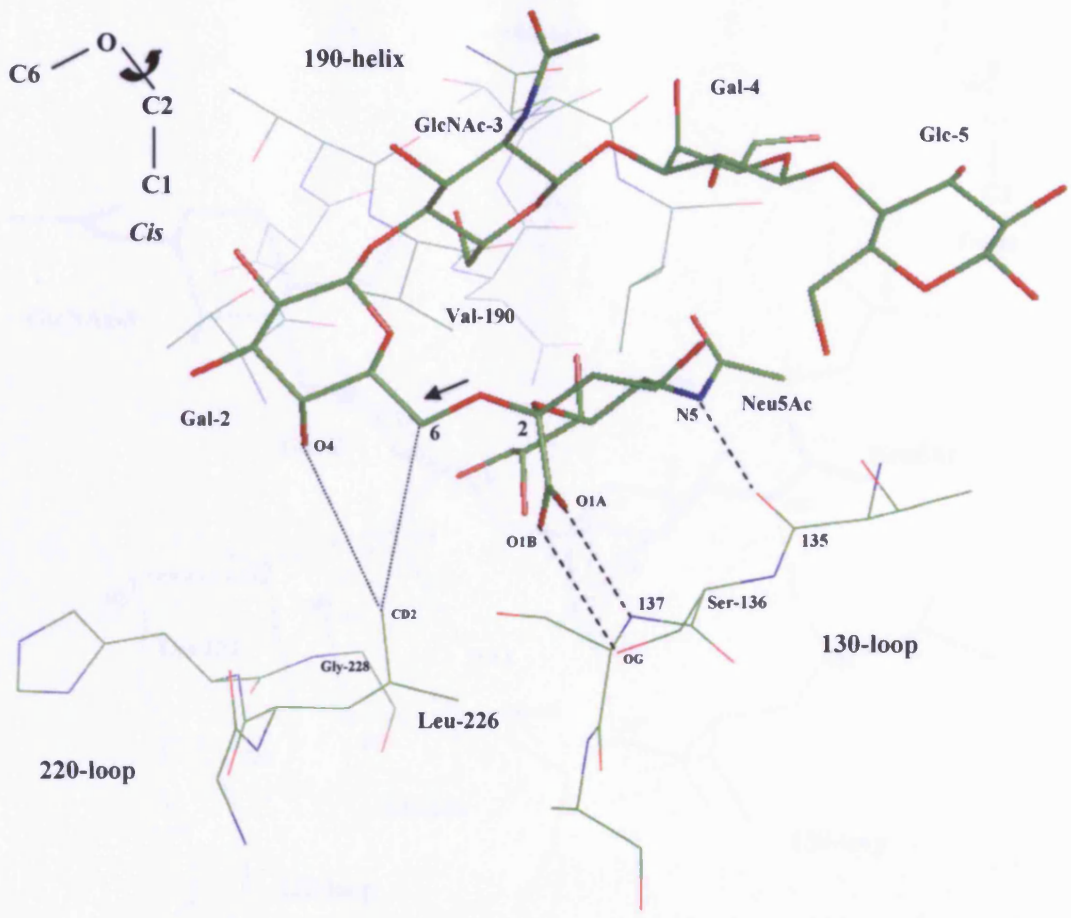


Figure 11B: Structure of H9 swine HA (Sw/Hong Kong/9/98) in complex with LSTc

The conformation of LSTc bound on the RBS of H9 swine HA is shown. The electron density allowed the location of all five saccharide residues of LSTc. Selected potential hydrogen bonds between analogue and H9 swine HA are shown as dashed lines. Atoms are coloured as: green for carbon, blue for nitrogen, and red for oxygen. Relevant residues of the HA RBS and atoms of LSTc and the HA RBS, namely those involved in the selected interatomic contacts, are specifically identified (see Table 11, Appendix IV). The three main secondary structural elements of the HA RBS - 130 loop, 190 helix and 220 loop - are also indicated. The arrow indicates the configuration around the glycosidic bond, as explained at the top left of the figure. The side chain of Leu226 makes van der Waals contacts with the Gal-2 of LSTc (dotted lines) (Ha et al., 2001). [The coordinates for H9 swine HA-LSTc were obtained from RCSB/PDB, accession code 1jsi]

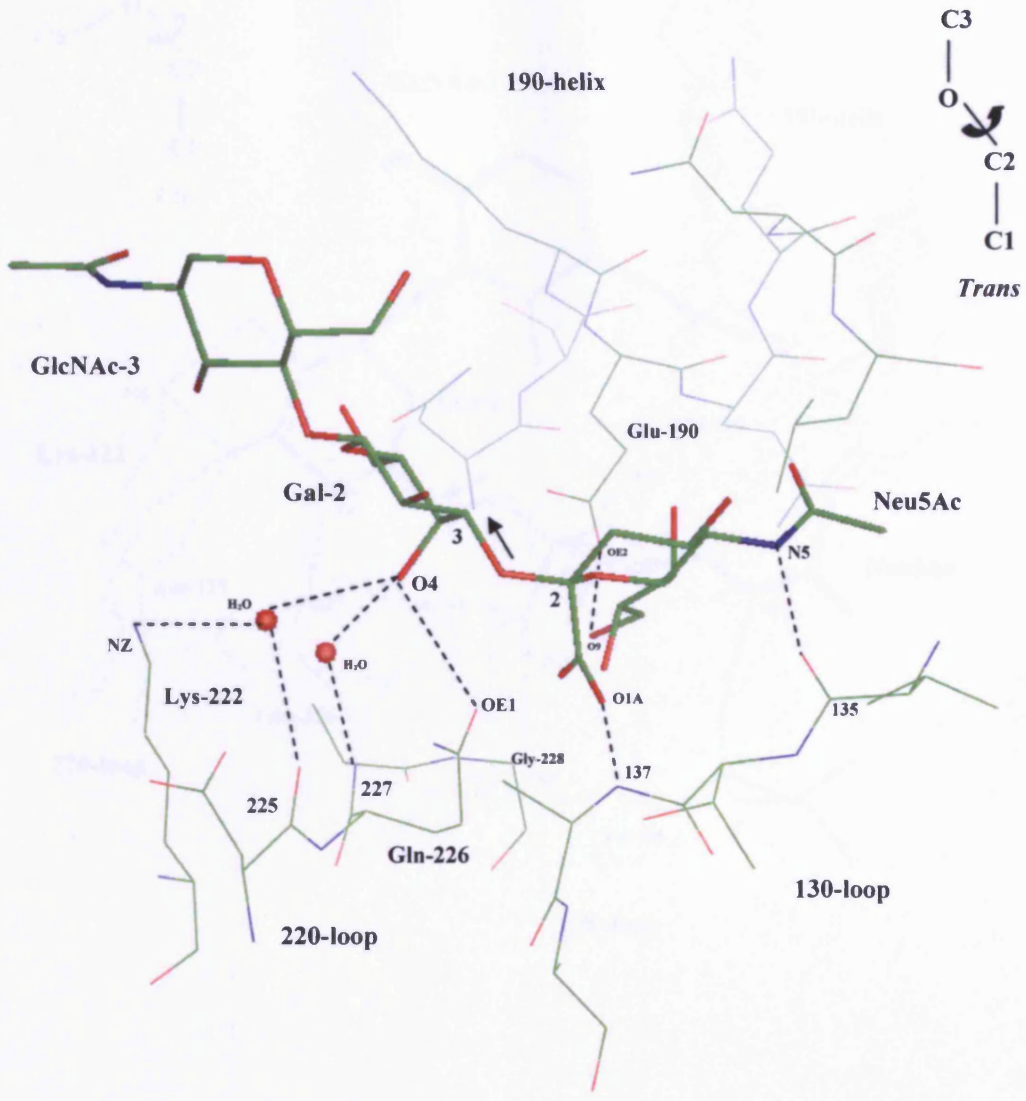


Figure 12A. Structure of H1 human HA (Puerto Rico/8/34) in complex with LSTa

The conformation of LSTa bound on the RBS of H1 human HA is shown. The electron density allowed the location of only the first three saccharide residues of LSTa. Selected potential hydrogen bonds between analogue and H1 human HA are shown as dashed lines. Atoms are coloured as: green for carbon, blue for nitrogen, and red for oxygen. Relevant residues of the HA RBS and atoms of LSTa and the HA RBS, namely those involved in the selected interatomic contacts, are specifically identified (see Table 10, Appendix IV). The three main secondary structural elements of the HA RBS - 130 loop, 190 helix and 220 loop - are also indicated. The arrow indicates the configuration around the glycosidic bond, as shown schematically at the top right of the figure (Gamblin et al., 2004). [The coordinates for H1 human HA-LSTa were obtained from RCSB/PDB, accession code 1rvx]

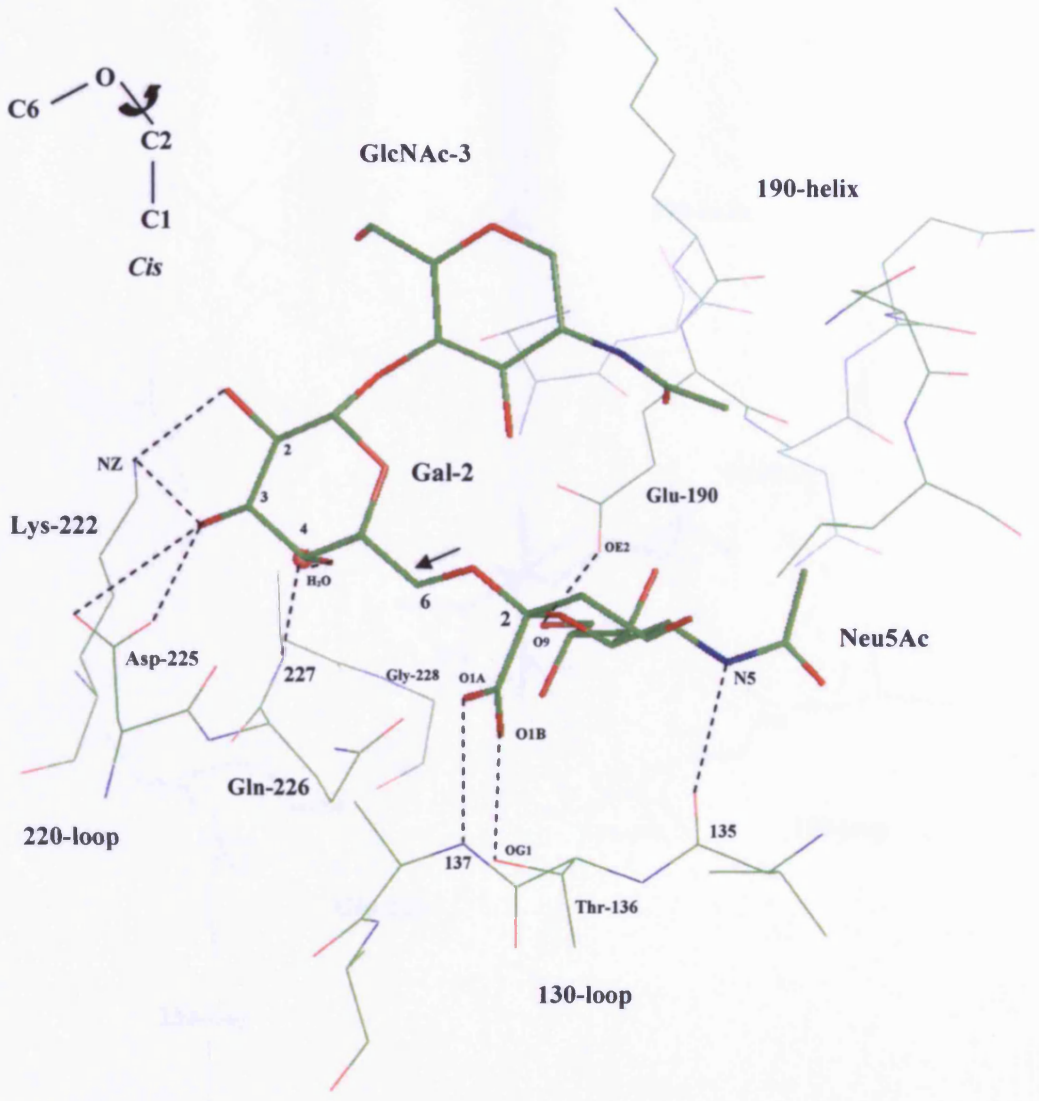


Figure 12B. Structure of H1 human HA (Puerto Rico/8/34) in complex with LSTc

The conformation of LSTc bound on the RBS of H1 human HA is shown. The electron density allowed the location of only the first three saccharide residues of LSTc. Selected potential hydrogen bonds between analogue and H1 human HA are shown as dashed lines. Atoms are coloured as: green for carbon, blue for nitrogen, and red for oxygen. Relevant residues of the HA RBS and atoms of LSTc and the HA RBS, namely those involved in the selected interatomic contacts, are specifically identified (see Table 11, Appendix IV). The three main secondary structural elements of the HA RBS - 130 loop, 190 helix and 220 loop - are also indicated. The arrow indicates the configuration around the glycosidic bond, as explained at the top left of the figure (Gamblin et al., 2004). [The coordinates for H1 human HA-LSTc were obtained from RCSB/PDB, accession code 1rvz]

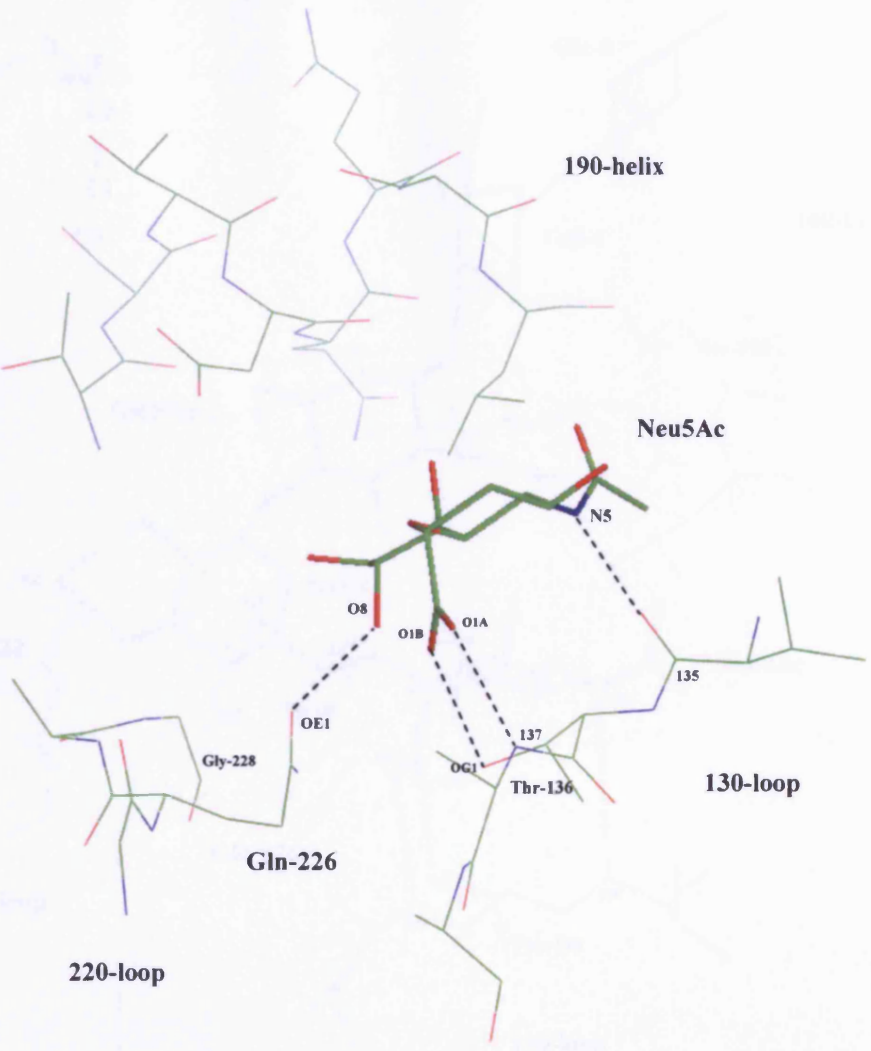


Figure 13A. Structure of H1 swine HA (Sw/Iowa/15/30) in complex with LSTa

Only the Sia of LSTa, bound on the RBS of H1 swine HA, could be fitted in the electron density. Selected potential hydrogen bonds between analogue and H1 swine HA are shown as dashed lines. Atoms are coloured as: green for carbon, blue for nitrogen, and red for oxygen. Relevant residues of the HA RBS and atoms of Sia and the HA RBS, namely those involved in the selected interatomic contacts, are specifically identified (see Table 10, Appendix IV). The three main secondary structural elements of the HA RBS - 130 loop, 190 helix and 220 loop - are also indicated (Gamblin et al., 2004). [The coordinates for H1 swine HA-LSTa were obtained from RCSB/PDB, accession code 1rvo]

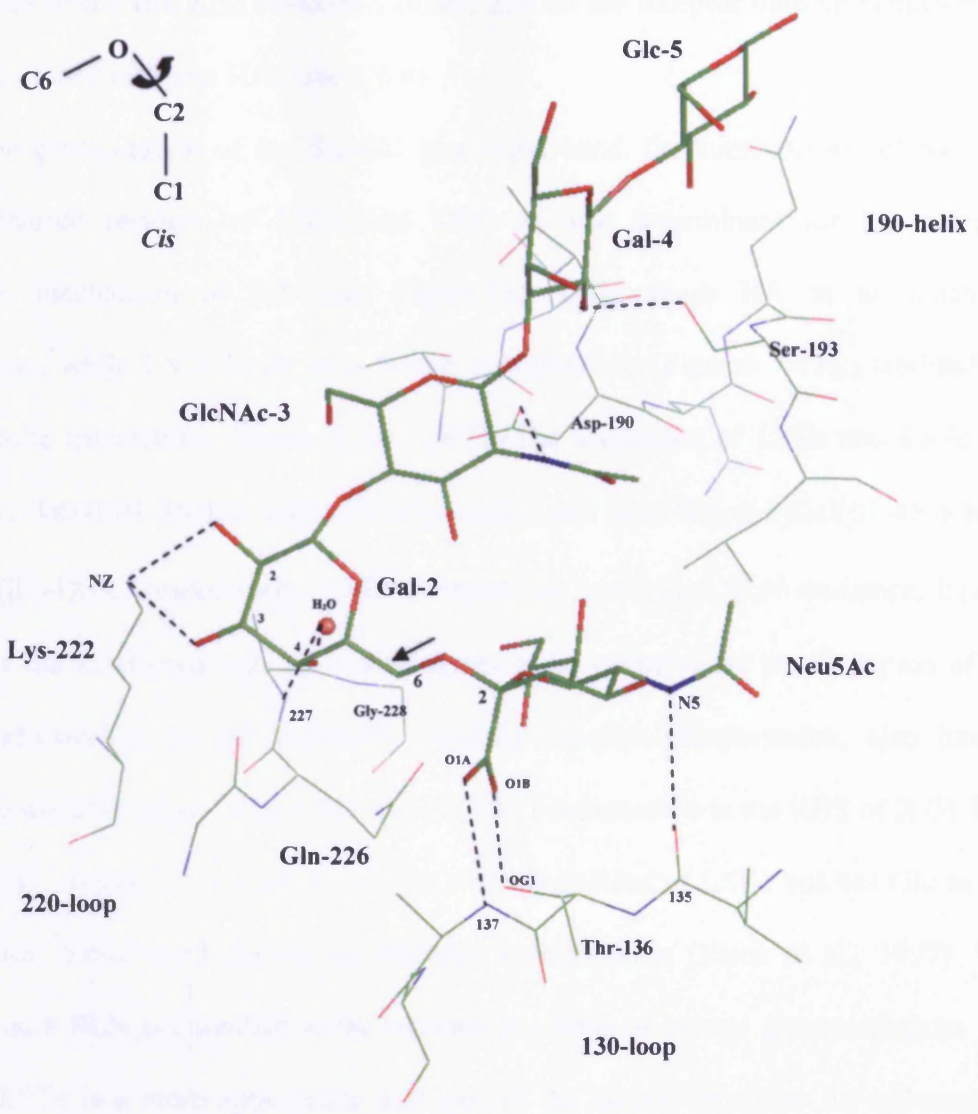


Figure 13B. Structure of H1 swine HA (Sw/Iowa/15/30) in complex with LSTc

The conformation of LSTc bound on the RBS of H1 swine HA is shown. The electron density allowed the location of all five saccharide residues of LSTc. Selected potential hydrogen bonds between analogue and H1 swine HA are shown as dashed lines. Atoms are coloured as: green for carbon, nitrogen (blue) and oxygen (red). Relevant residues of the HA RBS and atoms of LSTc and the HA RBS, namely those involved in the selected interatomic contacts, are specifically identified (see Table 11, Appendix IV). The three main secondary structural elements of the HA RBS - 130 loop, 190 helix and 220 loop - are also indicated. The arrow indicates the configuration around the glycosidic bond, as shown schematically at the top left of the figure (Gamblin et al., 2004). [The coordinates for H1 swine HA-LSTc were obtained from RCSB/PDB, accession code 1rvt]

origin in complex with LSTa and LSTc is shown in Table 1. This comparison focuses on the influence of the HA RBS residues 226 and 228 on the receptor binding behaviour of the just mentioned different HAs (see *1.6.6*).

Besides the configuration of the Sia-Gal glycosidic bond, the conformation of the first three saccharide residues of LSTa and LSTc is also determinant for the receptor recognition mechanism of HA (see Figure 16). LSTa binds HA in an extended conformation, while LSTc binds in a folded conformation (Figures 7-13), stabilised by intramolecular interactions (Eisen et al., 1997). The sequences of LSTa and LSTc are Neu5Ac(α2-3)Gal(β1-3)GlcNAc(β1-3)Gal(β1-4)Glc and Neu5Ac(α2-6)Gal(β1-4)GlcNAc (β1-3)Gal(β1-4)Glc, respectively. LSTc contains the terminal 6'SLN sequence, having GlcNAc as the third sugar residue, which seems to be essential for the formation of the folded conformation, as the unrelated inhibitor 3,6-disialyllactosamine, also having GlcNAc as the third sugar, adopts the same folded conformation in the RBS of X-31 HA, whereas 6'SL, which is identical to the fist two saccharides of LSTc but has Glc as the third residue, binds X-31 HA in an extended conformation (Eisen et al., 1997). The trisaccharide 6'SLN is abundant at the terminal positions of natural glycoconjugates and therefore LSTc is a more appropriate analogue of the natural receptors for influenza A virus (Matrosovich and Klenk, 2003; Eisen et al., 1997).

Examples of relevant HA-receptor analogue interactions in the structures of the complexes of LSTa and LSTc with H3 human HA, H3 avian HA, H5 avian HA, H7 avian HA, H9 swine HA, H1 swine HA and H1 human HA are shown in Figures 7-13, and a more detailed analysis in comparison with the LSTa-L226Q HA and LSTc-L226Q HA complexes is presented in Chapter 4.

HA	RBS residue 226	RBS residue 228	Linkage preference	$LSTa_{(\alpha2,3)}$ configuration	$LSTc_{(\alpha2,6)}$ configuration
H3 human	Leu	Ser	α2,6	*cis*	*cis*
H3 avian	Gln	Gly	α2,3	*trans*	*cis*
H5 avian	Gln	Gly	α2,3	*trans*	?
H7 avian	Gln	Gly	α2,3	*trans*	---
H9 swine	Leu	Gly	α2,6	*cis*	*cis*
H1 human	Gln	Gly	α2,6/α2,3	*trans*	*cis*
H1 swine	Gln	Gly	α2,6	?	*cis*

Table 1: Linkage type and configuration preferred by HAs of different subtypes and hosts of origin

Comparison of the Sia-Gal linkage type and configuration preferred by HAs of different subtypes and hosts of origin in relation to the RBS residues 226 and 228. The linkage preference of the different HAs for α2,3 or α2,6 linkages analysed in different binding assays (see *1.6.2*, *4.7.1.1* and *4.7.2.1*) and the Sia-Gal linkage configuration of LSTa and LSTc seen in the crystal structures of the HAs in complex with those receptor analogues are indicated (see Figures 7-13). The influence of the amino acid residue at the RBS positions 226 and 228 on receptor binding by HA is discussed in the text (see *1.6.6*, *4.8.1.2*, *4.8.1.3* and *4.8.4*). The H1 human HA binds equally to α2,6 and α2,3 linhages (dual preference) (see *4.8.1.1*). The H7 avian HA did not bind LSTc. The configuration of LSTa and LSTc in complex with H1 swine HA and H5 avian HA, respectively, is not known, since in the H1 swine HA-LSTa and H5 avian HA-LSTc structures only the Sia residue of the receptor analogue could be seen.

1.6.8 Correlation between linkage specificity and host of origin

1.6.8.1 Avian viruses

The primary natural reservoir of influenza A viruses are wild aquatic birds, which are hosts for all currently known 16 HA and 9 NA antigenic subtypes (Fouchier et al., 2005; reviewed in Webster et al., 1992; Alexander, 2000; Horimoto and Kawaoka, 2001). These viruses occasionally transmit to sea mammals, land-based poultry, horses, swine and humans, causing infections of different severities. The inter-species transmission of influenza viruses rarely gives rise to a virus adapted to the new host. The transmitted viruses usually have a poor fitness in the new host, only occasionally being able to efficiently replicate and transmit between individuals and continue to circulate for prolonged periods of time, forming a stable host-specific virus lineage. All known lineages of influenza A viruses identified in land-based birds and mammals derived from viruses of wild aquatic birds.

Wild ducks have a major role in the perpetuation of influenza viruses in nature (Webster et al., 1992). In ducks, the viruses mainly target cells of the intestinal tract mucosa, where they replicate, and do not lead to significant evidence of disease (Slemons and Easterday, 1977; Webster et al., 1978). The efficient transmission of these viruses through a fecal-oral route in the wild duck population (reviewed in Webster et al., 1992; Webster, 1997) together with the avirulent nature of the infection seem to have formed the conditions for the establishment of ducks as the natural reservoir for influenza viruses.

Avian viruses of different subtypes were reported to preferentially bind to α2-3-linked Sias than to α2-6-linked ones (Rogers and D'Souza, 1989; Nobusawa et al., 1991; Connor et al., 1994; Gambaryan et al., 1997; Matrosovich et al., 1997, 1999, 2000). Histochemical analysis of duck intestinal epithelial cells using linkage-specific lectins revealed the presence of Neu5Ac(α2-3)Gal-terminated sequences but no detectable expression of

Neu5Ac(α2-6)Gal (Ito et al., 2000). Confirming a lack of Neu5Ac(α2-6)Gal-terminated receptors in the duck intestine, human viruses were found not to bind to plasma membranes isolated from duck intestinal cells (Gambaryan et al., 2002). Therefore, the receptor specificity of duck viruses is linked to the predominance of Neu5Ac(α2-3)Gal-terminated oligosaccharides at the surface of the cells where the viruses replicate.

1.6.8.2 Swine viruses

In the laboratory and in natural conditions, swine is susceptible to infection by both avian and human influenza viruses (Hinshaw et al., 1981; Kida et al., 1994; reviewed in Scholtissek et al., 1998). The introduction of viruses from other hosts only rarely leads to the establishment of stable virus lineages, as observed in other mammals. H3N2 viruses, which have been circulating in humans since the 1968 pandemic originated the H3N2 human-like swine virus lineages.

Pigs are seen as 'mixing vessels', because they are easily infected with avian and human influenza viruses, forming the conditions for reassortment between the two differently originated viruses that can lead to the emergence of pandemic strains (Scholtissek et al., 1985). Only a few H3N2 swine viruses showed similar receptor binding specificity to human viruses (Higa et al., 1985; Suzuki et al., 1997). The presence of both Sia(α2-6)Gal and Sia(α2-3)Gal determinants can be detected on the surface of pig respiratory epithelia (Ito et al., 1998), which may explain the human virus-like specificity of pig viruses and the susceptibility of pigs to both avian and human viruses.

1.6.8.3 Human viruses

All human viruses have high affinity for receptor analogues with terminal Neu5Ac(α2-6)Gal moieties and usually bind weakly to Neu5Ac(α2-3)Gal-terminated receptors, as

shown in studies of the binding by many human influenza A (H1, H2 and H3) and B viruses to derivatised erythrocytes, gangliosides, sialylglycoproteins and sialylglycopolymers (Rogers and Paulson, 1983; Suzuki et al., 1985, 1986; Rogers and D'Souza, 1989; Xu et al., 1994; Connor et al., 1994; Ito et al., 1997; Gambaryan et al., 1997, 1999; Matrosovich et al., 2000). In most of these studies, α2-6-linked analogues were represented by the 6'SLN [Neu5Acα2-6Galβ1-4GlcNAc]. Presently, 6'SLN appears to be the most probable receptor determinant used by human influenza A and B viruses in nature, as it is abundant in natural sialylglycoconjugates and binds with high affinity to all human virus strains (Matrosovich and Klenk, 2003).

The ciliated epithelium of the human respiratory tract consists of several distinct cell types with different functions (Jeffery and Li, 1997). Neither the initial target cells for influenza virus nor the specific cell types supporting virus replication have been defined. In differentiated cultures of human airway epithelium, it was observed that during the course of a single-cycle infection, human viruses preferentially infect non-ciliated cells, whereas avian viruses, and an egg-adapted human virus variant with avian-virus-like receptor specificity, mainly infect ciliated cells (Matrosovich et al., 2004). This pattern correlated with the predominant location of α2-6-linked Sias on non-ciliated cells and of α2-3-linked Sias on ciliated cells. These results suggest that differences in replication and pathogenicity of human and avian viruses in humans may be related to the differential cellular tropism of the viruses.

1.6.9 Pandemic human influenza viruses and receptor specificity

Antigenic shift and antigenic drift are the major mechanisms of the evolution of influenza viruses in humans (reviewed in Kilbourne, 1987; Murphy and Webster, 1996; Cox and Subbarao, 2000). Antigenic shift occurs when a virus carrying HA or both HA and NA

genes from a different host emerges in the human population (reassortant viruses; see below). These new strains can cause a pandemic because humans lack immunity to the new HA and NA proteins, which are the two major influenza virus antigens (see *1.2*). The evolution of pandemic viruses continues by the accumulation of point mutations that result in progressive changes of the HA and NA amino acid sequences (antigenic drift). The three pandemics of the twentieth century - 'Spanish' flu (H1N1 virus, 1918), 'Asian' flu (H2N2 strain, 1957) and 'Hong Kong' flu (H3N2, 1968) - were caused by the transmission into humans of entire viral genomes or individual virus genes from non-human hosts. The 'Russian' flu in 1977 was caused by the reappearance of an H1N1 human virus that was genetically very similar to the virus circulating in the 1950s. The H2N2 lineage was substituted by the H3N2 in 1968, but descendents of the H3N2 and H1N1 pandemic viruses are still circulating in the human population together with type B viruses, which may have arisen by an ancient transmission of an avian virus to humans (Webster et al., 1992).

Two human pandemics in the last century were caused by reassortant viruses carrying HA (H3N2, 1968), or both HA and NA (H2N2, 1957), derived from avian influenza viruses (reviewed in Webster et al., 1992; Cox and Subbarao, 2000). Studies of receptor specificity revealed that the earliest virus isolates from these pandemics, in contrast with closely related avian viruses, bind strongly to 6'SLN-containing receptors and have a low affinity for α2-3-linked receptor determinants (Rogers and Paulson, 1983; Suzuki et al., 1986; Connor et al., 1994; Matrosovich et al., 2000). The pandemic of 1918 likely originated from an avian virus that was transmitted to humans as a whole (Webster et al., 1992; Taubenberger et al., 1997). Binding studies of recombinant viruses containing the HA and NA genes from the 1918 strain suggested a human virus receptor specificity for the 1918 pandemic viruses (Reid et al., 2003; Matrosovich et al., 2000; Kobasa et al., 2004).

Crystallographic studies of the 1918 HA, and two related H1 HAs in complex with LSTa and LSTc, also suggest the preference of the 1918 virus for human receptors (Gamblin et al., 2004). Therefore, the acquisition of a binding preference for 6'SLN in comparison with Neu5Ac(α2-3)Gal occurred relatively soon after the introduction of avian H1, H2 and H3 HAs in humans, at least by the time the viruses from these human epidemics were isolated. The H5N1 chicken viruses that directly infected humans in HK in 1997 (Claas et al., 1998; Subbarao et al., 1998) kept the avian receptor specificity (Matrosovich et al., 1999) but gave no indications of efficient human-to-human transmission. Thus, although avian viruses showing no significant modifications in their receptor binding preference can directly infect humans and even cause a fatal disease, changes in receptor specificity seem to be required for effective virus transmission between humans. H5N1 viruses of the present 'bird flu' outbreaks isolated from humans can bind both human (6'SLN) and avian (3'SLN) receptors. A few HA mutations, identified as mediating a shift to Siaα2,6Gal recognition, may be used as markers of pandemic potential (Yamada et al., 2006).

There are only a few amino acid substitutions in the HA of the earliest human isolates corresponding to H2 and H3 avian viruses (Bean et al., 1992; Connor et al., 1994; Klimov et al., 1996). These alterations included mutations at positions 226 (Gln to Leu) and 228 (Gly to Ser). Studies comparing the receptor specificity of laboratory-derived viruses with field virus isolates that differ at these HA positions indicate that the substitution Gln226→Leu226 is required for virus binding to α2,6-linked Sias (Rogers et al., 1983, 1985; Matrosovich et al., 1993, 1997, 2000; Vines et al., 1998) and that the additional mutation Gly228→Ser228 increases the binding affinity (Matrosovich et al., 2000). The HAs of H2 viruses isolated in 1957 that had Leu226 but maintained the avian-like Gly228 may represent the first intermediary HA in the adaptation of the avian H2 HA to humans (Klimov et al., 1996).

1.6.10 Changes in receptor specificity during egg-adaptation of human influenza viruses

Traditionally, influenza viruses are grown in the allantoic cavity of embryonated chicken eggs. However, it was a long ago recognised (Burnet and Bull, 1943) that egg-grown viruses differ considerably from the original virus isolates. They noticed that the original viruses isolated from human throat washes grew in the amniotic but not in the allantoic cavity of eggs. Viruses adapted to grow in the allantoic cavity gave rise to derivative viruses showing altered patterns of agglutination of chicken and guinea pig erythrocytes. The original to derivative changes were identified by different studies in the 1980s as the result of amino acid substitutions around the HA RBS (reviewed in Robertson, 1993). Several variants with different substitutions can be derived from a single human specimen, but usually only one or occasionally two substitutions are found in each variant. Viruses that belong to different HA types and subtypes acquire distinct substitutions, type- and subtype- and strain-specific, during egg adaptation. On the other hand, unlike egg-adapted variants, human viruses isolated in Madin-Darby canine kidney (MDCK) cell cultures are usually homogeneous with unaltered HA sequences (Katz et al., 1990; Robertson et al., 1990, 1991; Katz and Webster, 1992).

Cells of the egg amniotic membrane and MDCK cells express both Neu5Ac(α2-6)Gal and Neu5Ac(α2-3)Gal virus determinants, while cells of chicken embryo chorio-allantoic membrane (CAM) contain Neu5Ac(α2-3)Gal but not Neu5Ac(α2-6)Gal determinants (Ito et al., 1997). Gambaryan et al. (1997) demonstrated that non-egg adapted human influenza A and B viruses do not bind to Neu5Ac(α2-3)Gal-containing receptors, whereas all their egg-adapted variants acquire this ability. All egg-adapted viruses had increased affinity for receptors on CAM cells compared to their non-adapted parents and neither egg-adapted variant displayed a decreased affinity for inhibitors present in the allantoic fluid

(Gambaryan et al., 1999). All these results suggest that eggs select receptor-binding variants of human viruses that have increased affinity for Neu5Ac(α2-3)Gal-containing receptors on the target cells of CAM. By contrast, the cells of the amniotic membrane and MDCK cells appear to express sufficient amount of Neu5Ac(α2-6)Gal determinants to permit relatively unrestricted growth of human viruses.

The enhancement of HA atomic interactions with the Neu5Ac(α2-3)Gal moiety is one of the mechanisms proposed for the increased binding to CAM cells of viruses containing HAs with amino acid substitutions derived from egg-adaptation (Gambaryan et al., 1997; 1999).

Substitutions in the HA of egg-adapted H1, H3 and B influenza viruses, in addition to enhancing the affinity for Neu5Ac(α2-3)Gal-terminated receptors, also decrease the affinity of viruses for 6'SLN-containing receptors (Gambaryan et al., 1999). It was concluded that the decreased binding of egg-adapted human viruses to equine α_2-macroglobulin, which contains 6'SLN-terminated biantennary carbohydrates, was due to a decreased binding to the 6'SLN moieties associated with an increased binding to Neu5Ac(α2-3)Gal-terminated receptors. The selection of the X-31/HS virus variant containing the L226Q HA (used in this project) from the parent virus X-31 (Rogers et al., 1983a) would be connected with a decreased affinity of the L226Q HA for the 6'SLN moieties of the X-31 inhibitor α_2-macroglobulin and its enhanced affinity for Neu5Ac(α2-3)Gal-terminated sialosides in the chicken embryo chorio-allantoic cavity, which is devoided of Sia(α2-6)Gal receptor determinants (Ito et al., 1997). Evidence for the preferential binding of X-31/HS to 3'SL relative to 6'SLN in contrast with the preference of X-31 for 6'SLN in comparison with 3'SL was obtained in competitive binding assays of solid-phase immobilised viruses to soluble sialosides (Matrosovich et al., 1993, 1997).

1.7 HA-mediated Membrane Fusion

Receptor-binding by HA does not induce any significant conformational changes in HA (see above). This initial event of viral infection allows the virus to attach to the target cell surface and subsequently be captured by endocytic vesicles. The trigger for a fusion-inducing conformational change is the binding of one or more protons, as the pH of the endosome becomes progressively lower. Most strains of influenza A have a critical fusion pH of about 5.0-5.5, corresponding to the pH of a relatively late endosomal compartment. No particular titrating residue can be found as causing the conformational transition. A cooperative conformational rearrangement occurs immediately after a threshold of protein charge is obtained (Daniels et al., 1985).

1.7.1 HA structural changes required for membrane fusion

The crystal structure of the HA precursor HA_0 (Chen et al., 1998a) suggested that, after cleavage of the surface loop connecting HA_1 and HA_2, structural rearrangements occur in which the newly formed HA_2 N-terminus, presumably guided by electrostatic forces, moves into a charged cavity, burying ionisable residues implicated in the low pH-induced conformational change. This seems to be a priming event by which the membrane fusion potential of HA is activated. In the next cycle of infection, HA binds, through HA_1, to the terminal Sias of the cellular receptors and at the low endosomal pH, in the range of 5 to 6 for influenza A, the HA_1 subunits detrimerise, while maintaining their individual structure (Bizebard et al., 1995), and the HA_2 trimer undergoes extensive refolding (Bullough et al., 1994), leading to exposure of the fusion peptide and an irreversible conformational change of the molecule (Skehel et al., 1982). These fusion pH-induced HA structural changes underlie the membrane fusion process (Figure 14A). Using hydrophobic photolabels, it was shown that only HA_2 residues 1-22, corresponding to the fusion

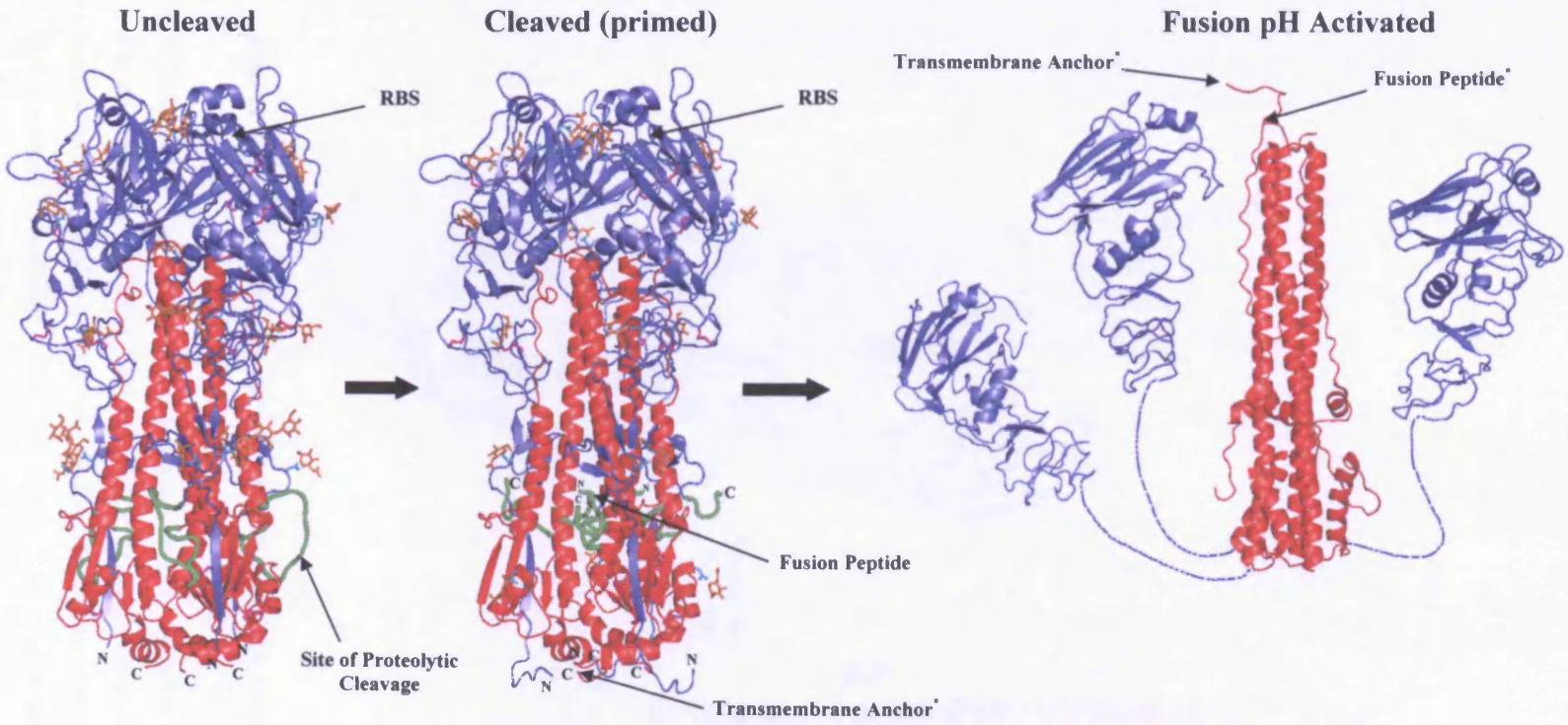


Figure 14. Molecular events of membrane fusion mediated by influenza HA

(A) HA conformational changes associated with its membrane fusion activity

The HA_0 precursor is cleaved at a loop (green) exposed on the surface of the molecule and the resulting fusion peptide, at the newly formed N-terminus of HA_2 (red), is buried in the interior of the trimer (priming event). At the fusion pH, the globular HA_1 (slate) domains at the membrane-distal tip of the molecule detrimerise and the HA_2 subunit refolds extensively to join fusion peptides and membrane anchors at the same tip of the new low pH-induced conformation (activation). The representation of HA at the fusion pH conformation corresponds to a hypothetical model, based on experimental indications (see text for details). In the latter representation, HA_1 is the X-31 HA_1 subunit (residues 44-309) and HA_2 the E. coli-expressed EHA_2 trimer (HA_2 residues 23-185; see *1.7.4*), and the connectors of the HA_1 subunits to the HA_2 trimer were drawn. The coordinates for HA_0, X-31 HA and EHA_2 were obtained from RCSB/PDB, accession codes 1ha0, 1hgf and 1qu1, respectively. Oligosaccharides and corresponding Asn side chains, and disulfide bonds are coloured as in Figure 3. *: potential location stemming from shown structure. [The figure was generated using the program Pymol (DeLano, 2006)]

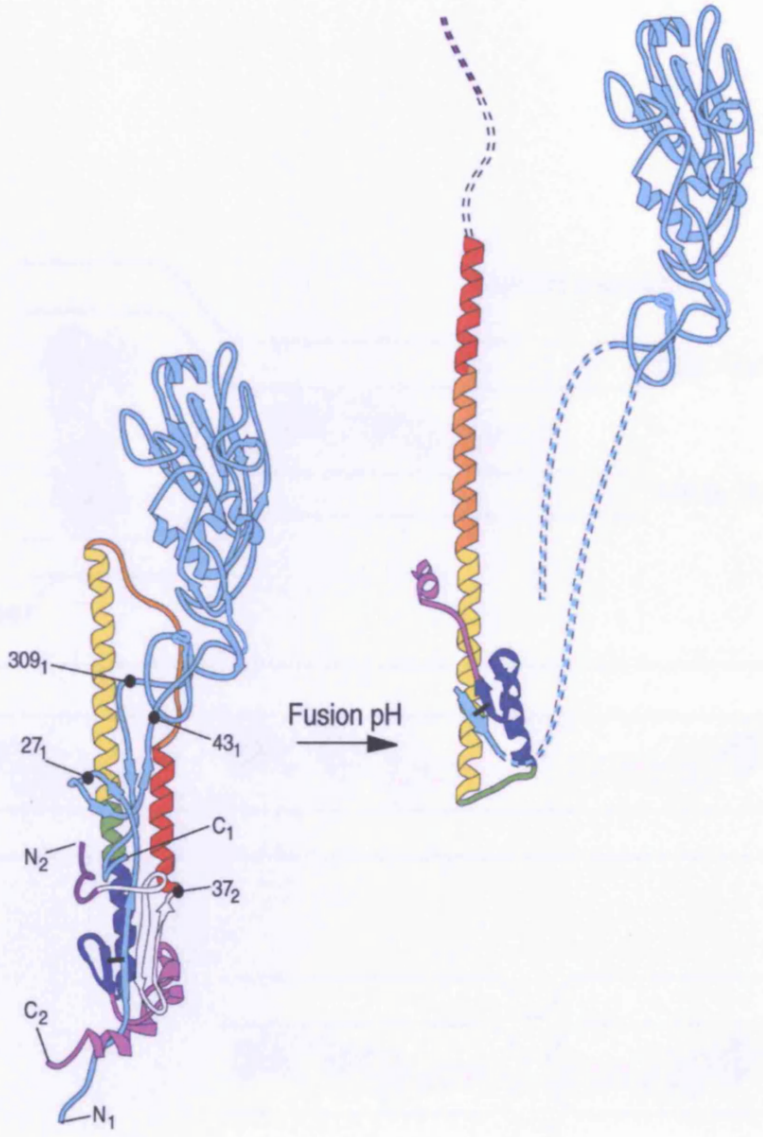


Figure 14B. Detail of the fusion pH-induced HA conformational change

Ribbon representation of the structural changes induced by fusion pH in each monomer of the HA molecule. The crystal structure of BHA (Wilson *et al.*, 1981) is shown on the left. HA_1 is shown in cyan and HA_2 in multicolour. HA_1 residue 27 and HA_2 residue 37 are recognised by trypsin and thermolysin, respectively. Before residue 43 and after residue 309, HA_1 is disordered in the determined crystal structure after fusion pH incubation (Bizebard *et al.*, 1995), shown in cyan on the right together with that of $TBHA_2$ in multicolour (Bullough *et al.*, 1994). The multicolour shows the details of the fusion pH-induced conformational change (see text). Discontinuous lines indicate regions of the molecule for which the crystal structure is not known in virus-derived protein.

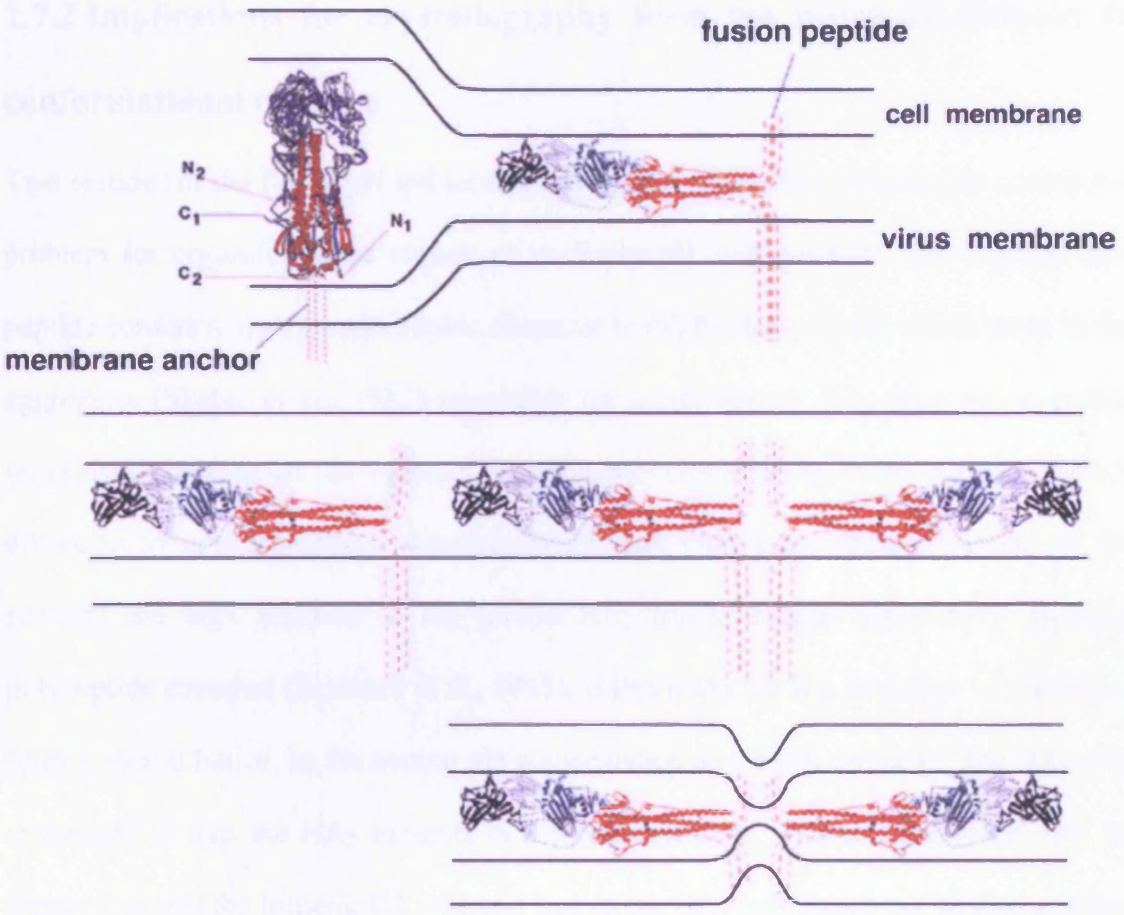


Figure 14C. Hypothetical mechanism for membrane fusion by influenza HA

After the priming and activation events described in Figure 14A, a number of HA trimers in the fusion pH-induced conformation form a ring-like complex, which mediates the formation of a fusion pore (see text for details). The HA_1 subunits are coloured in slate and the HA_2 trimer in red.

peptide, of HA in the fusion pH conformation were inserted into a membrane (Durrer et al., 1996). The covalent (Godley et al., 1992) or antibody-mediated (Barbey-Martin et al., 2002) cross-linking of the HA_1 subunits of the HA molecule prevents the fusion pH-induced conformational changes. The covalent cross-linking of the HA_1 subunits was also shown to abolish the membrane fusion activity of the HA.

1.7.2 Implications for crystallography from the fusion pH-induced HA conformational changes

Two features of the fusion pH-induced structural changes in the HA molecule constituted a problem for crystallographic studies of its fusion pH conformation. The exposed fusion peptide confers a highly hydrophobic character to the BHA molecule, which tends to form aggregates (Skehel et al., 1982) unsuitable for crystallisation. The aggregation problem was solved cleaving off the region of the molecule containing the fusion peptide. The other difficulty for crystallography stems from the fact that, upon detrimerisation, the HA_1 subunits are kept attached to the central HA_2 trimer through structurally disordered polypeptide stretches (Bizebard et al., 1995), which make the HA molecule too unstable to form a crystal lattice. In the neutral pH conformation of the HA molecule, the HA_1 trimer is thought to trap the HA_2 subunits in a metastable state. The endosomal pH acts as a trigger that sets the trimeric HA_2 subunit free of the HA_1 belt, assuming its lowest state of free energy. This feature of the fusion pH-induced conformational changes of HA indicates that the crystal structure of HA, in the neutral pH conformation, can only be known for the complete molecule and, in the fusion pH conformation, only for HA_2 or HA_1, separately.

1.7.3 Structural details of the fusion pH-induced HA conformational change

The details of the fusion pH-induced structural transition were determined by comparing the crystal structures of BHA and thermolysin-solubilised BHA_2 ($TBHA_2$) (Bullough et al., 1994). Indications that the BHA trimer had undergone a low pH-induced conformational change, exposing the fusion peptide, were the fact that BHA became susceptible to trypsin digestion and aggregated upon incubation at fusion pH (Skehel et al., 1982). $TBHA_2$ was derived from BHA aggregates (Ruigrok et al., 1988) by cleavage with trypsin, which removes the majority of HA_1, and then thermolysin, which removes a HA_2 N-terminal stretch containing the fusion peptide (Figure 15). The crystal structure of $TBHA_2$ indicated that three major fusion pH-induced refolding events occurred in the native HA molecule (Figure 14B). A loop and short α-helix (orange and red) in native HA became an extension of the central coiled coil in $TBHA_2$ (residues 76-105 extended to 38-105), presumably relocating the fusion peptide over 100Å from its previously buried position. The middle of the long α-helix (green) of native HA_2 unfolded to form a reverse turn, and the second half of the long α-helix (deep blue) jackknifed back, lying antiparallel to the first half and relocating the C-terminal residues by more than 100Å in relation to coiled-coil residues 76-105 (yellow), which remained unaltered in $TBHA_2$ relative to BHA. Residues 141-175 (purple), located C-terminally to a small β-sheet hairpin, appear to have been extruded from a compact association in the native HA_2 into a mostly extended structure, packing within the grooves of adjacent α-helices in the extended coiled-coil of $TBHA_2$.

1.7.4 *Escherichia coli*-expressed HA_2 molecules

Different constructions of HA_2 (Figure 15) expressed in *E. coli*, namely *E. coli*-expressed

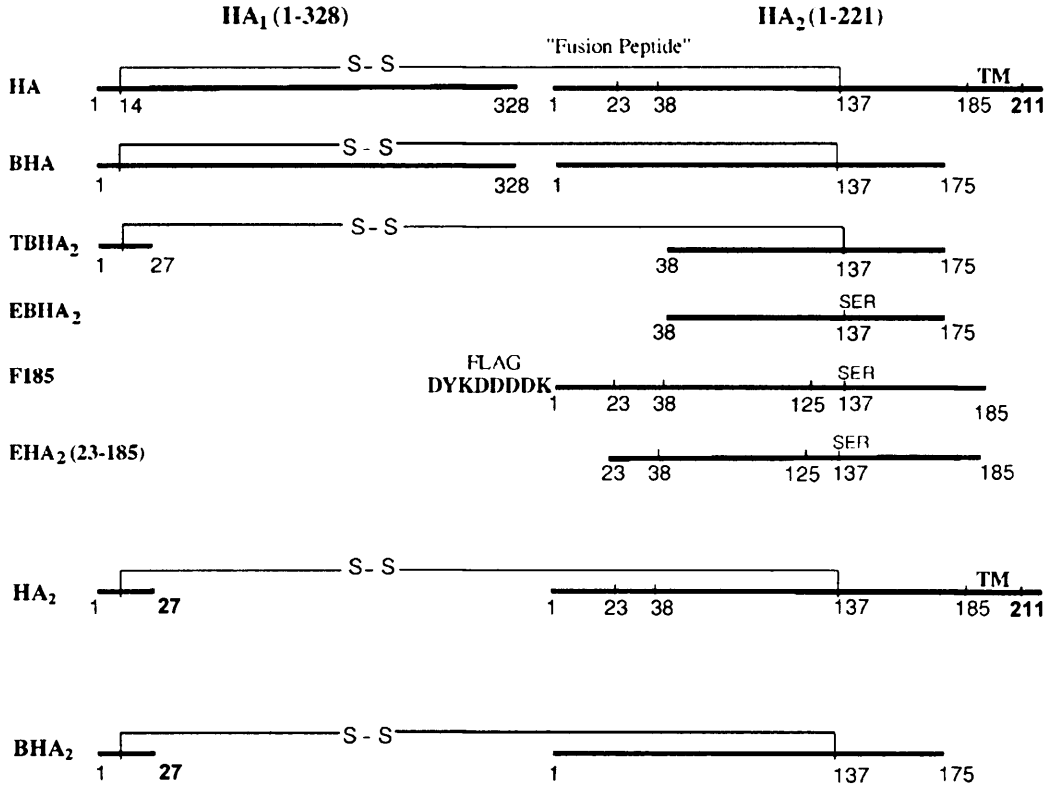


Figure 15. HA molecules previously studied or under study

Schematic of influenza virus-derived HA molecules, namely the ones used in the present study (HA, BHA, HA_2 and BHA_2) and previously studied (BHA, $TBHA_2$), and *E. coli*-expressed ($EBHA_2$, F185 and EHA_2). The crystal structure of $EBHA_2$ and F185 was not determined. TM: transmembrane anchor. S-S: interchain disulphide bond from HA_1 Cys 14 to HA_2 Cys 137. (Adapted from: Chen et al., 1999).

$TBHA_2$ ($EBHA_2$), residues 38-175, *E. coli*-expressed flag-tagged HA_2 (F185), residues 1-185 solubilised by a N-terminal hydrophilic tag, and *E. coli*-expressed soluble HA_2 (EHA_2), residues 23-185, spontaneously assembled in the fusion pH conformation (respectively: Chen et al., 1995; 1998b; 1999). This fact gave further indications that the viral native HA_2 is freed into its more stable state upon incubation at the fusion pH. The released free energy may be used as activation energy to overcome the unfavourable coulombic repulsive interactions as consequence of juxtaposition of the viral and endosomal membranes to be fused. The crystal structure of EHA_2 (Chen et al., 1999) confirmed the indications given by that of $TBHA_2$ that the C-terminal regions of HA_2 pack in a mostly extended arrangement against the inner trimeric coiled coil, forming a rod-shaped HA_2 molecule, and revealed the presence of a conserved structure terminating and capping the coiled coil, the N-cap domain, that may confer an extra stability to the molecule and ensure that transmembrane anchors and fusion peptides are firmly gathered together at the tip of the molecule.

1.7.5 Unknown structural aspects of the HA-mediated membrane fusion mechanism

Theoretically, the fusion of two lipid vesicles implies the interaction of their membranes to form pores through which the vesicle contents will interchange. Presumably, the pore formation starts by the fusion of the outer leaflets of membrane bilayers, producing a stalk structure. This defines a hemifusion event, as the transfer of contents between the fusing vesicles is not yet possible. The subsequent fusion of the inner monolayers completes the membrane fusion process resulting in the formation of a pore. Stepping from these concepts, the influence of lipid composition on fusion may reflect the necessary membrane curvature to form hemifusion and a complete fusion pore; the interaction of the fusion

peptide with the target membrane may facilitate the formation of stalks, leading to hemifusion; the membrane anchors, presumably by interaction with the fusion peptides in a trimer, are required to complete the pore; and the assembling of a multi-trimeric HA complex forming a ring-like fence around the hemifusion stalk may prevent the extension of hemifusion, favouring the pore formation (Figure 14C); also, the formation of a completely proteinaceous pore may precede the formation of a lipid pore. Experimental support and models for all these indications have been reported (reviewed in Skehel and Wiley, 2000; Cross et al., 2001; Tamm, 2003), but there is not a consensus that could define a general mechanism for a complete membrane fusion event.

The soluble molecules used in the previous crystallographic studies of HA, in both native or fusion pH-induced conformations, lacked exposed membrane-interacting regions, due to the problems associated with their aggregation-inducing character (reviewed in Skehel and Wiley, 2000). Therefore, the crystal structure of full-length HA in the native conformation containing transmembrane anchors or that of full-length HA_2 in the fusion pH-induced conformation containing fusion peptides or both fusion peptides and transmembrane anchors is not known. The presence of the membrane-interacting regions may have implications for the structure of the whole HA or HA_2 molecules (Figure 15). And the comparison of the structural details of the transmembrane anchors and fusion peptides in a trimer in the native or fusion pH-induced conformation would give indications for features like angle of entry, conformation inside the membrane and mediation of intra or inter-trimer contacts, helping to understand how the structure of the viral and endosomal membranes is disturbed, leading to their fusion.

In the native soluble HA molecule the highly conserved fusion peptide is in an extended conformation (Wilson et al., 1981). It is rich in glycines, suggesting that flexibility or potential for an unusual conformation of that region may be required for membrane fusion.

Structural spectrometric studies using synthetic fusion peptide analogues indicate about 45% α-helical content in liposomes. But the fact that these peptides depend on low pH for fusion and membrane interaction is in disagreement with observations that the pH of fusion by HA correlates with the pH of its conformational change (Skehel et al., 2001). In a construct made to solubilise the highly aggregating fusion peptide, this one assumed a monomeric V-shaped structure, forming a hydrophobic core, in both lipid bilayers and detergent micelles (Han et al., 2001). But this structure is subject to the caveats that it is fusion pH-dependent and the fusion peptide is not in the context of a trimeric molecule. Mutagenesis studies of intact HA molecules have defined critical residues for membrane fusion, suggesting a structure for the fusion peptide in a fusion active molecule by which it has a helical structure with residues orientated in such a way that if it forms a trimer, the relatively polar glycine residues may form the trimer interface and the large hydrophobic side chains would be on the surface of a coiled coil (Skehel et al., 2001). In a recent study, the HA fusion peptide showed a high conformational plasticity related to its adaptation to different environments (Vaccaro et al., 2005).

Even though the sequence of the transmembrane anchor is not highly conserved, the fact that glycosylphosphatidylinositol-anchored HA molecules are unable to mediate a complete fusion event (Skehel and Wiley, 2000) suggests an essential structural role for this region in the membrane fusion process. The transmembrane anchor comprises twenty-seven residues, which are sufficient to cross a lipid bilayer as a α-helix. This conformational arrangement would be favourable in the lipid environment. The study of HA molecules with truncated transmembrane anchors suggests the need of at least seventeen residues, enough to span the bilayer, for efficient fusion (Armstrong et al., 2000). Also, the formation of a triple-stranded coiled coil of helical transmembrane anchors inside the membrane would facilitate trimerisation of the C-terminal region of

HA_2, which may be important for the mechanism of fusion. HA transmembrane anchor synthetic peptides have a tendency to form oligomers in SDS micelles (Tatulian and Tamm, 2000), suggesting their trimerisation in the HA molecule.

The importance of the HA cytoplasmic tail, namely the palmitate residues, for membrane fusion is not clear. They were shown to be essential for pore flickering, a feature associated with early HA-mediated fusion pores, but not for the kinetics of fusion (Melikyan et al., 1997) and even infectious viral particles can be obtained for mutant influenza virus harbouring HA molecules not containing the cytoplasmic tail or palmitoylation (Jin et al., 1996).

The BHA_2 (HA_2 residues 1-175) fragment of HA in the fusion pH conformation, also prepared in this study (Figure 15), corresponds to a $TBHA_2$ molecule, whose crystal structure was previously determined (Bullough et al., 1994), having the remaining N-terminal region of HA_2, including the fusion peptide. The absence of the transmembrane anchor and palmitoylated cytoplasmic tail in BHA_2, could make the molecule less hydrophobic than HA_2, and thus less difficult to crystallise. It would reveal the structure of the trimeric fusion peptide of HA_2 in the fusion pH conformation. It would not, however, be able to form the N-cap motif found in the EHA_2 construct (Chen et al., 1999). The comparison of the BHA_2 and HA_2 structures could show structural changes in the trimetric fusion peptide induced by the transmembrane anchor.

1.8 Objectives

The crystal structure of the L226Q HA in complex with the receptor analogues LSTa and LSTc was determined to clarify the interactions the receptor makes with the RBS region that could contribute for the shift in viral receptor preference induced by the L226Q mutation. The structure of the L226Q HA in complex with LSTa revealed a dual character of the L226Q HA in the recognition of Sia(α2-3)Gal linkages, being able to simultaneously bind both the *cis* and *trans* conformers of LSTa, in two different RBSs. In the structure of the complex of the L226Q HA with LSTc, only the Sia moiety was observed, which suggests a low binding affinity of the L226Q HA to this receptor analogue. These data correlate with results from binding studies, which indicate a preference of the X-31/HS virus and L226Q BHA for α2-3-linked sialosides over α2-6-linked ones, in opposition to what is observed for the parent X-31 virus.

The structure of the native HA containing the transmembrane anchor or HA fragments in the low-pH conformation containing the fusion peptide or both the fusion peptide and transmembrane anchor is not known. These structures could reveal alterations induced by the presence of these membrane-interacting regions, which could even lead to the formation of inter-trimer complexes by the low pH-activated fragments of HA related to the formation of a fusion pore. For this reason, the purification of neutral pH full-length X-31 HA, containing the transmembrane anchor, low pH-treated full-length HA_2, having both the fusion peptide and transmembrane anchor, and low pH-treated BHA_2, which was cleaved from the virus with bromelain to remove the transmembrane anchor but still contains the fusion peptide (see schematic on Figure 15), was carried out, followed by crystallisation trials. These have not yet given any promising indication for further developments.

2. Materials and Methods

2.1 Protein Isolation, Purification and Crystallisation

2.1.1 Viruses and virus purification

X-31 (H3N2) recombinant influenza A virus (Kilbourne, 1969) was grown in the allantoic cavity of embryonated hens' eggs and purified as described previously (Skehel and Schild, 1971). X-31 contains the HA of the human influenza virus strain A/Aichi/2/68 (Verhoeyen et al., 1980) [HA gene sequence accession code at the National Center for Biotechnology Information (NCBI), nucleotide database: V01085], which is the prototype of the H3 antigenic subtype virus responsible for the 1968 HK flu pandemic. The virus variant X-31/HS containing the L226Q HA (Rogers et al., 1983a) was grown and purified following the procedures used for the X-31 virus.

The final pellet of purified virus was ressuspended in phosphate-buffered saline (PBS) [171 mM NaCl , 3.35 mM KCl , 10.1 mM Na_2HPO_4, 1.84 mM KH_2PO_4, pH 7.4], 0.01% azide or Tris saline buffer (10 mM Tris pH 8.0, 0.15 M NaCl, 0.01% azide).

2.1.2 Enzymes

Stock solutions of trypsin, trypsin inhibitor (TI) and bromelain were prepared as follows. Trypsin Type III, TPCK treated, from bovine pancreas (SIGMA) and Trypsin Inhibitor Type I-S from soybean (SIGMA) were dissolved in PBS to 1 mg/ml, divided into aliquots and kept at -20°C. Bromelain from pineapple stem (SIGMA) was dissolved in deionised water to 10 mg/ml and the resulting solution was centrifuged at 1000 *g* (bench centrifuge) for 10 min at 4°C. The pellet was discarded and the supernatant (SN) divided into aliquots and kept at -20°C. Small-scale TI columns were prepared by packing Soybean Trypsin

Inhibitor Gel (SIGMA).

2.1.3 Antibodies

Tissue culture SNs of the anti-NA monoclonal antibody (mAb) N154M (Ig G) at 1 mg/ml and ascites fluid of the anti-HA mAb HC73 (Ig G_{2a}) at 3 mg/ml were used as sources of mAbs for the immunoaffinity columns (see below).

2.1.4 Detergents

To keep the proteins soluble, detergents were used throughout the whole purification procedure and also in the crystallisation trials. Detergents or surfactants in general are amphipathic molecules with distinct hydrophilic (head group) and hydrophobic (tail) moieties. The detergents used to crystallise membrane proteins are generally simple nonionic or zwitterionic surfactants, which have only a single polar head group and a single unbranched, saturated alkyl tail. At dilute concentrations, surfactants are readily soluble in water as a monomeric species, but above the critical micellar concentration (CMC) the monomers begin to self-associate into micelles (Garavito, 1991). In the present case, the detergent molecules are expected to coat the exposed fusion peptides and transmembrane anchors and also the palmitate residues of the cytoplasmic tail, solubilising the molecules of the neutral pH conformation of HA, and the fusion pH conformation of HA_2 and BHA_2.

The detergents used were the nonionic Polyoxyethylene 10 Lauryl ether ($C_{12}E_{10}$) from SIGMA [linear structural formula (LSF): $CH_3(CH_2)_{10}CH_2(OCH_2CH_2)_{10}OH$; MW = 627.0; CMC = 0.2 mM] and the zwitterionic Lauryldimethylamine oxide ($C_{12}DAO$) from Fluka [LSF: $CH_3(CH_2)_{10}CH_2NO(CH_3)_2$; MW = 229.4; CMC = 1-3 mM]. Both detergents are included in the commercially available crystallisation detergent screens from Hampton.

Concentrated stock solutions were prepared for each detergent, 20%$_{w/v}$ $C_{12}E_{10}$ and 10% $_{w/v}$ C_{12}DAO. The stock solution of $C_{12}E_{10}$ was prepared dissolving the detergent in 10 mM Tris pH 8.0 and that of C_{12}DAO by dilution of the commercial aqueous solution with deionised water. Both $C_{12}E_{10}$ and C_{12}DAO solutions were kept at room temperature (RT).

2.1.5 Determination of protein concentration

Protein concentration was estimated from measured A^{280} values. The concentration of mAbs was calculated assuming that A^{280} = 1 corresponds to 0.8 mg/ml (Millar, 1997). Concentration reference values (C_{ref}) corresponding to A^{280} = 1 were calculated for the proteins under study using the relationship C_{ref} (mg/ml)= MW × 1 ml × A^{280} / (No. Trp residues × 5700 + No. Tyr residues × 1300). The estimated values of protein concentration probably have a maximum accuracy of 10% (Ruigrok et al., 1986). Theoretical MWs obtained by submitting the protein sequences to the ExPasy server were used to calculate the reference values. These are: 0.680 (HA), 0.781 (BHA), 0.653 (HA_2), and 0.922 (BHA_2).

2.1.6 Isolation of viral envelope proteins by detergent solubilisation

Viral suspensions (see *2.1.1*) at 5 mg/ml were incubated in the solubilisation buffer (2% $C_{12}E_{10}$ in PBS, 0.01% azide) for 1 hour at 4°C. After incubation, the solubilisation mixture was ultracentrifuged at 200000 *g* for 15 minutes at 4°C, in an Optima™ MAX-E ultracentrifuge, using a MLA-80 rotor and 8 ml tubes (Beckman). The obtained pellet (virus cores) was resuspended in PBS, 0.01% azide and kept together with the SN (viral membrane proteins, mainly HA) at 4°C.

2.1.7 Isolation of viral envelope proteins by proteolysis

Viral suspensions (see *2.1.1*) at 5 mg/ml were incubated with the cysteine protease bromelain, at a ratio of $2{:}1_{w/w}$ virus:bromelain, in a solution of 50 mM 2-mercaptoethanol (2-ME) in PBS, 0.01% azide, at 37°C. Three-step digestions were performed, one of 90 minutes followed by two overnight (ON) steps. At the end of each step, the digestion mixture was ultracentrifuged at 200000 *g* for 15 minutes at 4°C, in an Optima™ MAX-E ultracentrifuge, using a MLA-80 rotor and 8 ml tubes (Beckman). The resulting pellet was resuspended and reincubated in the conditions given above for the subsequent digestion and the SN run in a linear sucrose gradient (see *2.1.8*) or subject to ion exchange (IE) chromatography (see *2.1.13*). Bromelain digestions were stopped by both the latter procedures, which separated bromelain and 2-ME from the BHA. The final pellet was resuspended in PBS, 0.01% azide and kept at 4°C.

2.1.8 Sucrose density gradient (SDG) ultracentrifugation

Linear sucrose gradients were prepared at RT by mixing in a linear gradient maker diluted solutions of a stock solution of sucrose, using the same buffers of the stock solution. Concentrated stock solutions were 60% sucrose in PBS or 0.1% C_{12}DAO, 10 mM Tris pH 8.0, 0.15 M NaCl. After adding the protein solutions (the SNs from *2.1.6* and *2.1.7*) onto 10 ml sucrose gradients, at a maximum load of 0.3 mg protein/ml gradient (V_{max}=1.5 ml), the obtained mixtures were ultracentrifuged at 150000 *g* ON at 4°C, in a an Optima™ L-90K ultracentrifuge, using SW-41 rotors and 12 ml tubes (Beckman). Fractions of 1 ml were then collected from the gradients and kept at 4°C. Some gradients were performed with a cushion (1 ml) of sucrose stock solution at the bottom.

2.1.9 Dialysis

From the collected fractions (V=1 ml) of the ultracentrifuged sucrose gradients (see, above, *2.1.8*), fractions N^{os} 4, 5 and 6 (corresponding to the middle of the gradient, see Figure 21), which are enriched in HA or BHA (see *3.1.3*), were selected. These chosen fractions were pooled, introduced into a 10 kDa MW cut-off (MWCO) Slide –A-Lyzer® dialysis cassette (PIERCE) and then incubated at 4°C with the dialysis buffer [10 mM Tris pH 8.0 (for BHA); 0.1 % C_{12}DAO, 10 mM Tris pH 8.0 (for HA)]. The dialysis was performed in one quick initial step of 5 hours followed by three steps of one day. Sometimes the samples to be dialysed were concentrated by filtration (diluting the sample with dialysis buffer, to avoid concentration of the sucrose) in order to fit the volume into the dialysis cassette.

2.1.10 Protein concentration

Protein concentration by centrifugation was carried out using the Vivaspin Concentrators (Sartorius) Vivaspin 500 and Vivaspin 20, according to manufacturer's the instructions, in an Eppendorf 5402 benchtop centrifuge (Eppendorf) and a GS-6R benchtop centrifuge (Beckman), respectively. Concentrators containing 30 kDa or 10 kDa MWCO membranes were used for samples with or without the detergent C_{12}DAO, respectively.

Protein concentration by filtration was performed in an AMICON® concentrator, connected to a N_2 cylinder, using 30 kDa or 10 kDa MWCO OMEGA® membranes (FILTRON) for samples with or without C_{12}DAO, respectively. The stirring of the samples was kept slow to not induce foaming.

2.1.11 Immunoaffinity chromatography

Three different chromatography columns were prepared: uncrosslinked anti-NA,

crosslinked anti-NA and crosslinked anti-HA. The anti-NA columns were prepared by binding the mAb to a prepacked HiTrap rProtein A FF column (Pharmacia), and the anti-HA by binding the mAb to a packed Protein A Sepharose® 4 Fast Flow gel (Pharmacia). Crosslinking of the anti-NA mAb to protein A was only possible using the HiTrap column. Each mAb stock solution was diluted to 0.5 mg/ml with PBS, passed through a 0.45 μm filter and then passed through the column pre-equilibrated with 20 vols PBS. The column was washed with PBS and kept at 4°C. Crosslinking of mAb to protein A, to prevent detergent-induced mAb leakage from the column, was carried out using the imidoester cross-linker dimethyl pimelimidate (DMP) from PIERCE, following a procedure adapted from a described method (Schneider et al., 1982). Briefly, 20 vols 30 mM DMP, 0.2 M triethanolamine pH 8.3 were recirculated for 45 min by the column pre-equilibrated to the triethanolamine buffer and the reaction was quenched with the same volume of 0.3 M glycine pH 8.3. At the end the column was washed with PBS, 0.01% azide and kept at 4°C. Protein samples were passed slowly through the prepared immunoaffinity columns pre-equilibrated to the protein buffer. The columns were then washed and kept at 4°C in the protein buffer containing 0.01% azide.

2.1.12 NA activity assay

Protein samples were incubated with fetuin, a source of Sia, and assayed for free Sia following a described procedure (Warren, 1959). Protein samples, 100 μl, were mixed, in serial dilutions, with 50 μl of fetuin solution and incubated ON at 37°C. Periodate solution, 50 μl, was added and the mixture shaken thoroughly and incubated at RT for 20 min. Arsenite solution, 250 μl, was added and the mixture shaken until a yellow-brown colour disappeared. Freshly prepared thiobarbituric acid solution, 500 μl, was added and the mixture shaken and heated in a boiling water bath for 15 min. A relative neuraminic

activity was estimated by comparing the colour intensity of the final product obtained for the different samples, using the sample buffers as negative controls. The following solutions were used: 0.2 M sodium metaperiodate in 9 M phosphoric acid; 10% sodium arsenite in 0.5 M sodium sulfate, $3\%_{v/v}$ concentrated sulphuric acid; 1.2% thiobarbituric acid in 1 M sodium sulfate. The fetuin solution was prepared as follows. Fetuin was precipitated by mixing, in equal volumes, foetal calf serum with a solution of saturated ammonium sulfate at 4°C. After centrifugation at 2000 *g* for 1 hour at 4°C, the pellet was resuspended, washed and dialysed ON with distilled water. The dialysed solution was transferred to a separating funnel and left for 7 days at 4°C. The lower phase was divided into aliquots and kept at -20°C. The fetuin working solution was prepared diluting three times with PBS.

2.1.13 Ion-exchange (IE) chromatography

The IE chromatography was performed using the Vivapure Quaternary ammonium (*Q*) spin columns (Sartorius) Vivapure *Q* Mini and Vivapure *Q* Maxi in an Eppendorf 5402 benchtop centrifuge (Eppendorf) and a GS-6R benchtop centrifuge (Beckman), respectively, according to the manufacturer's instructions. The IE membranes included in the spin columns contain the strong basic anion exchanger *Q* ($-CH_2-N^+-(CH_3)_3Cl^-$). At the pH 8.0 of all the buffers used here, the proteins have a net negative charge and will bind to the *Q* membrane, under no-salt conditions. The procedure was carried out in four sequential steps: equilibration, binding, washing and elution. In the first step the column was equilibrated with the same buffer of the protein solution (loading buffer). The protein solution was then passed through the column to allow the protein to bind to the IE membrane. In the third step, the columns were washed with a buffer having the same composition of the elution buffer, but no salt (washing buffer). The protein was eluted

from the spin column with a buffer containing a concentration of NaCl high enough to practically elute all the protein molecules from the IE membrane (elution buffer). Salt gradients were carried out to establish the optimal concentration of salt for the final elution of the protein sample.

2.1.14 Trypsin digestions

Trypsin digestions were performed in detergent, either C_{12}DAO or $C_{12}E_{10}$. Before being incubated with the trypsin, the protein solutions were subject to a low pH treatment as follows. To lower the pH, aliquots of 0.1 M citric acid or 0.1 M citrate pH 3.5 were added to the protein solutions until the desired pH was obtained. After incubation at RT, the pH was readjusted for the trypsin reaction by adding aliquots of 1 M Tris pH 8.0 or 0.1 M citrate base. All buffer solutions used to adjust the pH contained detergent, either 0.1% C_{12}DAO or 2% $C_{12}E_{10}$. The trypsin reaction was set up adding the trypsin and incubating at RT. The reaction was stopped by adding an equal weight of TI or, for digestions in C_{12}DAO, by passing the final reaction mixture through a TI column pre-washed with 0.1% C_{12}DAO, 10 mM Tris pH 8.0.

2.1.15 Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was performed by the discontinuous Laemmli system (Laemmli, 1970) using 12% polyacrylamide gels containing 4 M urea, 0.1 % SDS. Samples for gel analysis were boiled in 8 M urea, 2 % SDS (non-reduced conditions) or 8 M urea, 2% SDS, 0.2% 2-ME (reduced conditions).

2.1.16 Matrix assisted laser desorption/ionisation (MALDI) spectrometry

The proteins to be analysed by MALDI spectrometry (Karas et al., 1985) were subject to SDS-PAGE, stained with SimplyBlue SafeStain (Invitrogen) and processed in-gel using a modification of a described method (Blackman et al., 1998). Briefly, after washing out the Coomassie stain from the excised gel bands, the proteins were reduced and alkylated with dithiothreitol (DTT) and iodoacetamide, respectively. The bands were then washed with 20 mM ammonium bicarbonate, 50% acetonitrile, dehydrated with 100% acetonitrile, dried and incubated with 2 mg/ml trypsin (sequencing grade, Promega) in 5 mM ammonium bicarbonate for 20 hours at 32°C. The MALDI spectrometry analysis and peptide mass fingerprinting (Pappin, 1993) were carried out by S. Howell at the Division of Molecular Structure, MRC-NIMR, London. The peptide mass fingerprints obtained were searched against the nonredundant protein database placed in the public domain by NCBI, using the program MASCOT (Perkins, 1999). This program allowed the identification of the protein contained in the gel bands by matching the experimental tryptic peptide mass values, determined by MALDI spectrometry, with theoretical tryptic peptide mass values of the protein sequences in the NCBI database. Two clear highest scores of match were obtained for the full-length HA_2 amino acid sequence (HA_2 residues 1-221) of the influenza virus strain A/NT/60/68/29c (score: 94%) and the BHA_2 amino acid sequence (HA_2 residues 1-175) of X-31 virus (score: 91%) (protein sequence accession codes at NCBI, protein database: CAA24291 and 2VIUB, respectively). The A/NT/60/68/29c HA_2 sequence contains the E132D mutation in comparison with the A/Aichi/2/68 HA_2 sequence of the X-31 virus used in this study (see *2.1.1*). This mutation does not affect the trypsin cleavage of HA.

2.1.17 Protein Crystallisation

2.1.17.1 General Protein Crystallisation

The crystallisation of proteins initially involves the determination of their purity. Even though the crystallisation of a protein from a mixture of proteins can sometimes be obtained a highly pure protein sample is generally required for successful crystallisation. Therefore, the crystallisation process can be seen as a form of protein purification. A suitable solvent is necessary to keep the protein soluble until crystallisation conditions are induced. The solvent is usually a water-buffer solution, and occasionally organic solvents may be added. Membrane proteins, due to their high hydrophobicity, very often require the addition of detergents.

The spontaneous formation of very small protein aggregates that can nucleate the deposition of protein is required for the subsequent growth of crystals. Conditions of supersaturation are the most appropriate to induce the formation of crystallisation nuclei. The adjustment of supersaturation, which affects the number of nuclei formed and rate of crystal growth is necessary to obtain suitable crystals for crystallography. Too many nuclei would lead to too many small crystals, and highly ordered crystals are usually obtained when they are allowed to grow slowly.

The precipitation of a protein leading to its crystallisation can be induced by several different methods. Usually, the addition of well-defined precipitants that remove water from the protein solution increases the concentration of the protein, leading to its precipitation. The most common precipitants are salts (salting-out effect) and polyethyleneglycols (PEGs). A salting-in effect may also be used to induce the crystallisation of proteins showing low solubility in pure water. Adding salt to the water makes the protein soluble, and by removing the salt, the precipitation of the protein is provoqued. Salting-in may be regarded as an electrostatic effect and protein-protein

electrostatic forces are affected by the presence of organic solvents or changes in pH. Another factor affecting protein crystallisation such as temperature may be used to increase the strength of protein-protein hydrophobic forces (Drenth, 1999).

Membrane proteins are water-insoluble and difficult to crystallise. Usually, detergents are used to solubilise these proteins in an aqueous solution and the procedures followed for water-soluble proteins are used (Michel, 1990; Sowadski, 1994).

2.1.17.2 Crystallisation trials

Crystallisation conditions for X-31 HA, HA_2 and BHA_2, at 18°C, were set up using both the hanging drop vapour diffusion (HDVD) and microbatch (MB) techniques (Unge, 1999). MB conditions were set up automatically using an IMPAX micro-dispenser (Douglas Instruments). Drops were prepared by mixing equal volumes of protein solution and precipitant solution: 0.7 µl, for HDVD, and 0.3 µl, for MB. The following commercially available crystallisation screens were used: MemStart and MemSys (Iwata, 2003), Nuclear Receptor–Ligand Binding Domain (NR-LBD), Structure Screen 1 (Jancarik and Kim, 1991), and Structure Screen 2 (Jancarik and Kim, 1991; Cudney et al., 1994), from Molecular Dimensions; Crystal Screen (Jancarik and Kim, 1991) and Crystal Screen 2 (Jancarik and Kim, 1991; Cudney et al., 1994), from Hampton; and Wizard I and Wizard II, from Emerald BioSystems. A PEG 5000 monomethylether (MME) salt screen (grid screen at pH 7.5), designed by L. F. Haire (Division of Molecular Structure, MRC-NIMR, London) and the described Core Screen (conditions 1-48) (Page et al., 2003) were also used. The MemSys screen was only set up by HDVD. Crystallisation drops were regularly examined under a light stereomicroscope. To check if the protein concentration of the samples used in crystallisation trials (see *3.1.6*, for HA, and *3.2.7.2*, for HA_2 and BHA_2) was appropriate for screening, pre-crystallisation test drops were

prepared by mixing 1 µl of each protein solution with 1 µl of each Crystal Screen (Hampton) conditions N° 4 (2.0 M ammonium sulfate, 0.1 M Tris-HCl pH 8.5) and N° 6 [30%$_{w/v}$ PEG 4000, 0.2 M $MgCl_2$ $(H_2O)_6$, 0.1 M Tris-HCl pH 8.5]. In general, the drops showed the formation of light granular precipitates and were therefore considered in conditions for crystallisation trials.

In the crystallisation trials of the purified L226Q BHA (see *4.1*), several commercially available screens and screens prepared in the laboratory (see below) were used.

The majority of the trials for the crystallisation of the L226Q BHA were set up at 18°C by the HDVD technique (Unge, 1999). Drops were prepared by mixing 0.5 µl of protein solution and 0.5 µl of reservoir solution. Initially, the crystallisation conditions of the L226Q BHA (1.34-1.38 M tri-sodium citrate at pH 7.5) previously described (Weis et al., 1988) were screened. No crystals were obtained in these trials. Therefore, the crystallisation conditions of the L226Q BHA were extensively screened using the above-mentioned NR-LBD and PEG 5000 MME screens, and the following commercially available screens: The Classics (Jancarik and Kim, 1991; Cudney et al., 1994), from Nextal; JBScreen 1-10, from Jena Bioscience; PEG/Ion and Additive Screen 1, from Hampton; Clear Strategy Screen I and Clear Strategy Screen II (Dauter et al., 2000), and PACT premier, from Molecular Dimensions. Several promising conditions of these trials were optimised but no good diffracting crystals were obtained.

Grid screens of the crystallisation conditions for L226Q BHA were also carried out, at 18°C by the sitting drop vapour diffusion technique (Unge, 1999), using an automated system (Oryx 1-6 robot, Douglas Instruments), which has set up 200 nl drops by mixing 100 nl of protein solution and 100 nl of precipitant solution. These screens were available in the laboratory (Division of Molecular Structure, MRC-NIMR, London) and trials were

carried out screening the concentration and pH values in solutions of tri-ammonium citrate, K citrate, $NaKPO_4$ and ammonium sulfate. Tri-ammonium citrate and ammonium sulfate were also screened in the presence of 1% PEG 400. The crystals used to determine the structure of L226Q BHA in complex with LSTa and LSTc (Figure 30) were obtained in one condition of these screens, namely in 2.1 M tri-ammonium citrate pH 7.5 (see *2.2.1*).

2.2 Determination of the X-ray Crystal Structure of L226Q HA in Complex with the Receptor Analogues LSTa and LSTc

2.2.1 Soaking of L226Q HA crystals with LSTa and LSTc for X-ray analysis

Crystals of L226Q BHA, grown in 2.1 M tri-ammonium citrate pH 7.5 (see 2.1.17.2), were captured in synthetic fibre loops and washed by transfer into drops of the reservoir solution. This enabled the removal of any uncrystallised HA and contaminant NA remnants, contained in the mother liquor, from the crystals. Attempts at obtaining complexes of L226Q BHA with LSTa and LSTc, by soaking crystals with the receptor analogues, together with preparation for flash cooling, were carried out by transferring the washed crystals, mounted in a fibre loop, into a drop of the precipitant solution containing cryoprotectant (glycerol) and either LSTa or LSTc (soaking buffer: 2.1 M tri-ammonium citrate pH 7.5, $10\%_{v/v}$ glycerol, 1.5 mM LSTa or LSTc), incubating in the drop for different times. After this procedure, the crystals were picked up from the soaking buffer using a fibre loop and flash cooled by plunging into liquid nitrogen.

The pentasaccharides LSTa [Neu5Ac(α2-3)Gal(β1-3)GlcNAc(β1-3)Gal(β1-4)Glc] (MW = 998.9) and LSTc [Neu5Ac(α2-6)Gal(β1-4)GlcNAc(β1-3)Gal(β1-4)Glc] (MW = 998.9)

were purchased from NBS Biologicals. The structural chemical formulas are given on Figure 16.

2.2.2 Collection of X-ray diffraction images

Preliminary X-ray diffraction collection tests on the crystals were done at 100 K (in a nitrogen gas stream) on the in-house rotating Cu anode generator (Rigaku-MSC RU200) coupled to a RaxisIIc detector. Full collection of final data sets was carried out at the Daresbury Synchrotron Radiation Source (SRS). Usually, a full data set is collected up to the resolution limit, defined by the ratio signal/noise ≈ 2. Optimal data can be collected at a synchrotron source, due to the high intensity of the X-ray beam, which is tunable, allowing the utilisation of short wave lengths. The loops containing the cryo-frozen crystals (see, above, *2.2.1*) were mounted on the goniometer head and centered in the nitrogen gas stream. Once a satisfactory diffraction image was obtained, the crystal was rotated 90° and a second image was taken. These two preliminary diffraction images were processed with the program Denzo (Otwinowski and Minor, 1997) to determine the crystal orientation (see, below, *2.2.3.1*), and then the collection strategy, to maximise the data coverage, was defined using the program ROTGEN (Campbell, 1995), which outputs total angles φ of crystal oscillation range for different levels of data coverage.

2.2.3 X-ray data processing

The HKL suite (Otwinowski and Minor, 1997), version 1.96.3, including the programs *XdisplayF*, for visualization of the diffraction pattern, Denzo, for data reduction and integration, and SCALEPACK, for merging and scaling of the intensities given by Denzo, was used for the processing of the collected data. The procedures of these programs are

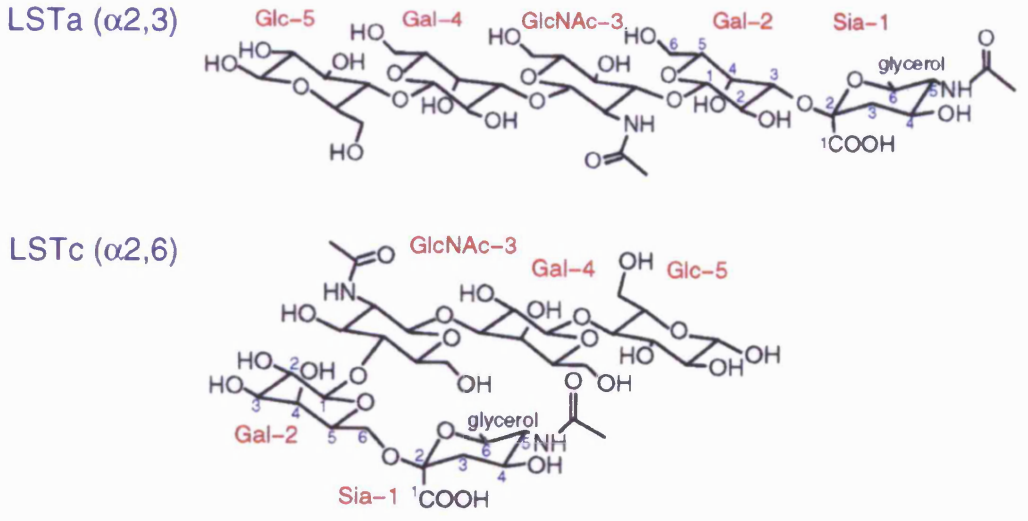


Figure 16. Structural chemical formulas of LSTa and LSTc

Structural chemical formulas of the pentasaccharides LSTa and LSTc, showing the adopted extended and folded conformations, respectively, on the binding to HA. The numbering of the pyranose cyclic carbon atoms of Sia and Gal-2 is indicated. [Edited from Ha et al., 2001]

briefly described below and an outline of all data processing steps, including additional programs to prepare the data for the subsequent procedures of molecular replacement (see *2.2.4.2.1*) and model refinement (see *2.2.4.4*) is shown in Figure 17.

2.2.3.1 Data reduction and integration

The autoindexing of selected spots of a diffraction image carried out by the program Denzo (Otwinowski and Minor, 1997) allows the determination of the lattice type and unit cell dimensions of the crystal, together with the definition of the orientation of the crystal relative to the X-ray beam. Subsequently, Denzo performs the reduction of raw X-ray diffraction data, by integration of each of the diffraction images, producing a final file containing a list of the Miller indices (*h,k,l*), unscaled intensities (I) and associated standard deviation (σ) corresponding to each reflection.

Initially, the crystal and detector orientation parameters are determined for a single image. The intrinsic crystal parameters describe the lattice type and unit cell dimensions; the crystal orientation parameters, the orientation of the reciprocal lattice to the spindle, beam and vertical axes of the camera; the detector and X-ray parameters, the wavelength, crystal to detector distance, the precise coordinates of the direct beam, detector missetting angles and internal scanner alignment parameters. Subsequently, the above-mentioned parameters are refined using all frames collected.

The display program *XdisplayF* (Otwinowski and Minor, 1997) allows interactive monitoring of the progress of the Denzo refinement procedures. This is done by displaying the match between the predicted lattice points, resulting from space group and unit cell selection, with the observed pattern as parameters are refined. Also, after each refinement cycle, Denzo lists statistics indicators, which give a quantitative indication of the progress of the refinement. An important statistic is the value of χ^2, which is a weighted

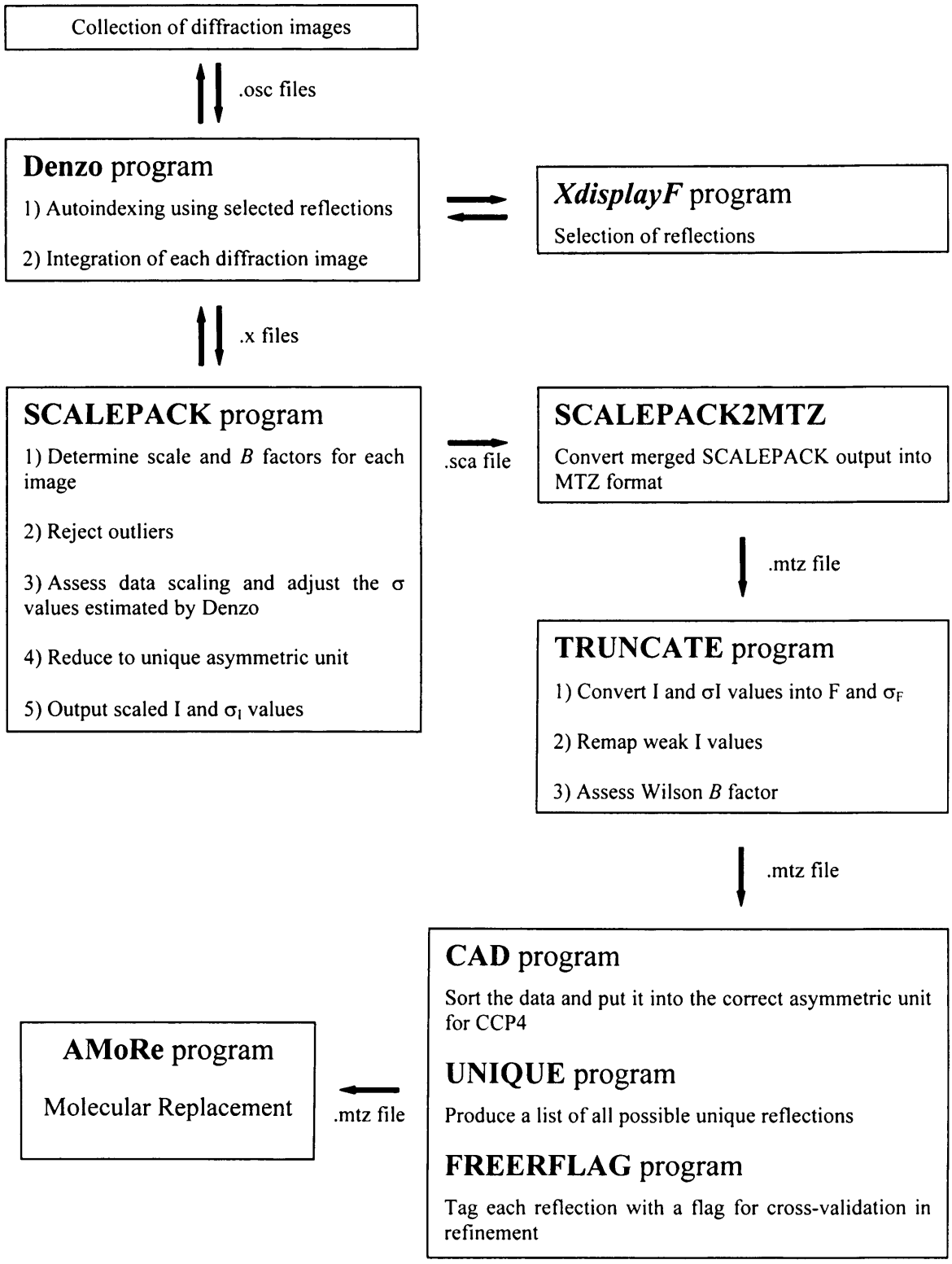


Figure 17. Flow chart showing the different steps of data processing

The programs Denzo, *XdisplayF* and SCALEPACK are included in the HKL suite (Otwinowski and Minor, 1997). The programs SCALEPACK2MTZ, TRUNCATE, CAD, UNIQUE and FREERFLAG are all part of the CCP4 program suite (Collaborative Computational Project, Number 4, 1994). The program AMoRe (Navaza, 2001), also implemented in the CCP4 program suite, was used here for molecular replacement (see *2.2.4.2.1*).

measure of the positional error of the predicted reflections. The decrease of χ^2 is an indication of successful refinement and, at the final refinement cycle, χ^2 ought to be about 1.0. The data can then be scaled using the companion program SCALEPACK (Otwinowski and Minor, 1997).

The number of spots selected for autoindexing, monitored with *XdisplayF*, can be adjusted by manipulating the values for the mosaicity and I/σ(I) (signal/noise). The aim is to have all real reflections matched by predicted reflections, but not having too many predictions. The value for the mosaicity obtained after refinement in Denzo (Otwinowski and Minor, 1997) was optimised by SCALEPACK (Otwinowski and Minor, 1997). Then the diffraction images were reprocessed with Denzo using the optimised value for the mosaicity.

2.2.3.2 Merging of reduced X-ray diffraction data

The program SCALEPACK calculates, for each frame of integrated reflections, the scale and *B* factors (see *2.2.4.3*, for *B* value), which are then used for a global refinement (postrefinement) of certain crystal parameters, namely the unit cell dimensions, mosaicity and misseting angles. The merging and scaling of diffraction images allows the estimation of the errors affecting the values of intensity (I) of each reflection. The program uses the equivalence between symmetry-related reflections to refine a global scaling solution. At the end, accurate unit cell values are obtained and the quality of the data can be assessed by the formula:

$$R_{symm} = \frac{\Sigma_h(I_h - \langle I_h \rangle)}{\Sigma \langle I_h \rangle}$$

where I_h is the scaled intensity for each reflection h, $\langle I_h \rangle$ is the weighted intensity of all observations of the reflection h and the summation includes all recorded measurements. The output file is then converted to MTZ format, and the square roots of the intensities are recorded as structure factor amplitudes (CCP4 suite).

2.2.4 Structure determination

2.2.4.1 The phase problem

In a diffraction experiment, the intensities of waves reflected from planes (denoted by *hkl*) in the crystal are measured. The amplitude of the wave $|F_{hkl}|$ is proportional to the square root of the intensity ($I_{h,k,l}$) measured on the detector. The calculation of an electron density map requires estimates of the phases of the waves scattered by the crystal. To calculate the electron density at a position (*xyz*) in the unit cell of a crystal, the following summation over all the *hkl* planes has to be carried out:

$$\rho(xyz) = 1/V \, \Sigma_{hkl} \, |F_{hkl}| \exp[-2\pi i(hx + ky + lz) + i\alpha_{hkl}]$$

where V is the volume of the unit cell (in Å) and α_{hkl} is the phase angle associated with the structure factor amplitude $|F_{hkl}|$, and x, y, and z are fractional coordinates in the unit cell. The amplitudes can be measured, but the phases are lost in the experiment and must be determined indirectly. This is the phase problem of crystal structure analysis.

2.2.4.2 Phasing methods

To determine the phases it is necessary to use phasing methods, namely isomorphous replacement, anomalous scattering, molecular replacement or on occasion direct methods

(Taylor G., 2003). In isomorphous replacement, heavy atoms are introduced in the crystal to create intensity changes that can be used to deduce their position in the unit cell. The heavy atom positions can then be used to calculate scattering factors, and with the Harker construction, determine the phases (Figure 18A). The anomalous scattering method is based on the presence in the protein of sufficient anomalous scattering atoms, which absorb X-rays at a specific wavelength and can generate reasonable anomalous differences (in general, stronger for heavier atoms) that can be used to locate these atoms. With the atomic coordinate parameters it is possible, again, to construct phase determination diagrams (Figure 18B and C). Direct methods are based on the positivity and atomicity of electron density that leads to phase relationships between normalized structure factors from which the phases can be determined. Once the phases of some reflections are known, a process of phase calculation is initiated leading to phase values for all reflections. The very high resolution required by these methods, higher than 1.2 Å, limits their application to protein crystallography. The molecular replacement method, used in this study, will be discussed below.

2.2.4.2.1 Molecular replacement

When a homology model is available, molecular replacement can be successful, using methods first described by Michael Rossman and David Blow (1962). Patterson search techniques are used for the first two steps: the orientation of the model in the new unit cell is determined by finding when the calculated Patterson function best matches the observed one (Rotation Function) and then the correctly orientated model is translated relative to the origin of the new unit cell (Translation Function). At the end, the positional parameters are usually optimised by rigid-body refinement.

The Patterson function is a map of vectors between the atoms within the unit cell. These

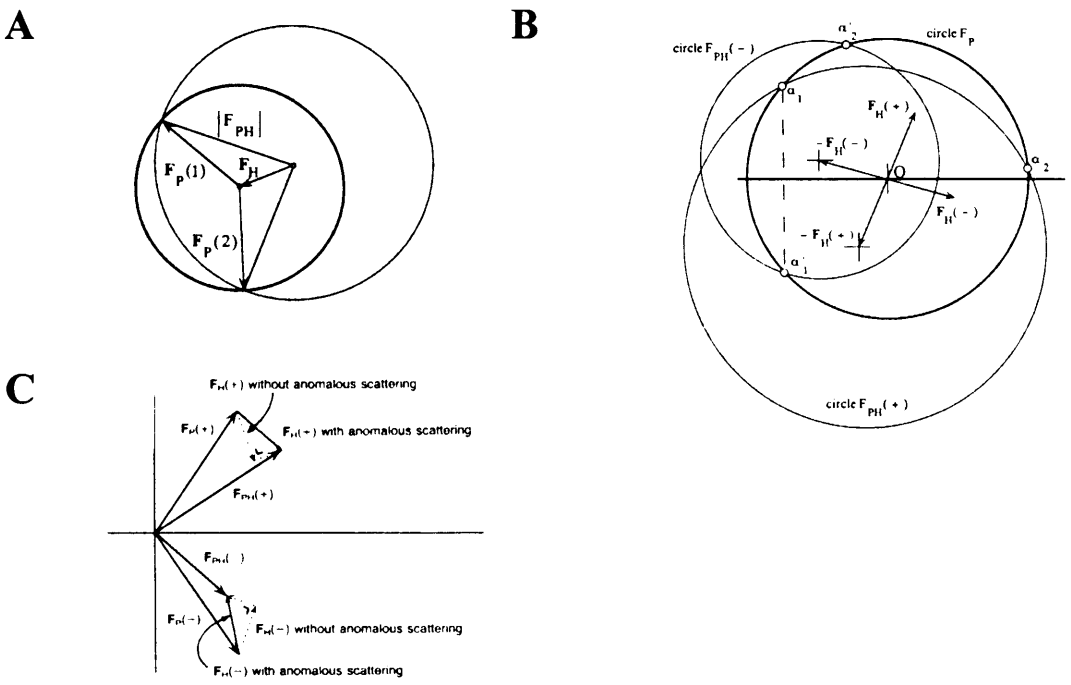


Figure 18. Protein phase determination by isomorphous replacement and anomalous scattering

(A) Harker diagram for phase determination by isomorphous replacement. The two possibilities for the phase angle of the protein (P), defined by the vectors $F_P(1)$ and $F_P(2)$, are obtained at the intersections of the circles with radii corresponding to the experimentally determined $|F_{PH}|$ and $|F_P|$ from the heavy atom derivative (PH) and native data sets. The heavy atom (H) coordinates are found by application of the Patterson function. Theoretically, using a second heavy atom derivative, a single value for the protein phase angle can be obtained.

(B) Harker diagram for phase determination by anomalous scattering. The phase angle (α) can be found with the construction obtained from three circles having radii F_P, $F_{PH}(+)$ and $F_{PH}(-)$. When anomalous scattering is present, Friedel's law, which would imply that $F_{PH}(+) = F_{PH}(-)$ for the reflections $h\ k\ l$ and $\bar{h}\ \bar{k}\ \bar{l}$ (Bijvoet pairs), is no longer obeyed (see below). For this reason, $F_H(+)$ and $F_H(-)$ are not symmetric in relation to the horizontal axis. As the native data has not an anomalous component, the phase angles α_P will be symmetric with respect to the horizontal axis (α_1 and α'_1).

(C) Argand diagram showing the anomalous scattering effect for a Bijvoet pair of reflections. When an anomalous scatterer is present, $F_H(+)$ and $F_H(-)$ are no longer symmetric in relation to the horizontal axis (see above). In this case, the intensity of a reflection $h\ k\ l$ is different from that of $\bar{h}\ \bar{k}\ \bar{l}$.

vectors can be divided into intramolecular or self-vectors, and intermolecular or cross-vectors, which arise between pairs of atoms in neighbouring molecules. The Patterson function can be calculated using the formula:

$$P_{uvw} = 1/V \, \Sigma_{hkl} \, |F_{hkl}|^2 \cos [2\pi (hu + kv + lw)]$$

where u, v, and w are coordinates in Patterson space, referring to the same axis and unit cell, but not related to the atomic coordinates x, y, z; and $|F_{hkl}|^2$ is derived from the directly measured intensities.

For a Rotation Function, the number of self-vectors is maximised, whilst minimising the cross vectors. This is done by placing the molecule in a large P1 cell 25-30 Å larger than the unit cell under study, and selecting a radius of search 75-80% the diameter of the molecule. This can be a complex procedure when the shape of a molecule is not spherical and the appropriate radius, which will include most of the self-vectors without too many cross vectors, is evaluated by trial and error.

The program AMoRe (Navaza, 2001), as implemented in the CCP4 program suite, was used here. This program includes routines to run a complete molecular replacement. The package allows for multiple inputs to the individual programs. The outputs are lists of potential solutions (Navaza, 1994). Two preliminary programs must be run to organise the data into a suitable format: SORTING, which sorts, packs and assesses the quality of the measured data, and TABLING, which calculates the continuous Fourier coefficients corresponding to the search model. During this step, the model coordinates are translated and rotated to fit a minimal box, which just contains the model. The ROTING program is then used to calculate the cross-Rotation Function, displaying a list of Eulerian angles (α, β, γ). The subsequent TRAING program is used to compute the Translation Function.

Finally, FITING is used to refine the positional parameters of the potential solutions by rigid body refinement. The program LSQKAB (CCP4 suite) applied the rotation and translation parameters obtained in AMoRe to the coordinates derived from TABLING.

2.2.4.2.1.1 Search model for the structure of L226Q HA

The 2.9 Å crystal structure of L226Q HA from its complex with 3'SL [Neu5Acα2-3Galβ1-4Glc] (coordinates accession code at RCSB/PDB: 4hmg) was used as the search model for molecular replacement. This corresponds to an improved model, after refinement by simulated annealing (Weis et al., 1990), of the original structure of the L226Q HA in complex with 3'SL (Weis et al., 1988). The electron density maps only allowed the placement of the Sia of 3'SL, in all three RBSs. Therefore, the structure of the asialo portion of 3'SL in complex with the L226Q HA, namely the configuration of the Sia-Gal α2-3 linkage, was not determined.

Here, in the initial model generated by LSQKAB after running AMoRE (see, above, *2.2.4.2.1*), the coordinates of the Sia of the 3'SL were replaced by models of the LSTa and LSTc receptor analogues during the refinement procedures (see, below, *2.2.4.4.1*).

2.2.4.3 Calculation of structure factors

From the structural model obtained by molecular replacement, which contains the coordinates *xyz* of the atoms of the structure, the theoretical structure factors (F_c) can be calculated by:

$$F_c = \Sigma_i f_i [\cos 2\pi (hxi + kyi + lzi) + i \sin 2\pi (hxi + kyi + lzi)]$$

In the above equation, f_i is the scattering factor (in electrons), which depends on the type of atom; it is modulated by the thermal factor, B (assuming isotropic and harmonic vibration):

$$B = 8\pi^2 \times \overline{u^2}$$

where $\overline{u^2}$ is the mean square deviation (*m.s.d.*) of the atomic vibration.

The vector F_c contains the phase information. A Fourier synthesis can then be performed, using the values of the amplitudes, $|F_o|$, and phases, α_{calc}, for the calculation of the electron density map corresponding to the structure.

2.2.4.4 Refinement

Refinement is necessary to improve the accuracy of the structure obtained by the molecular replacement calculation. Refinement leads to the improvement of the reliability of atomic parameters, and therefore α_{calc}, minimising the differences between the observed, F_{obs}, and calculated, F_{calc}, structure factors. Especially in the present case, the accurate determination of the interactions between protein and ligand is required to define the specific binding of LSTa and LSTc on the RBS of the L226Q HA.

In protein crystallography, refinement is generally based on the principle of maximum likelihood. This is a more general method of refinement than the least squares method, which can be applied if the observations have a Gaussian distribution. Refinement implies the variation of atomic positional parameters (x, y, z, B_{factor}) so that the values of F_{calc} get as close as possible to those of the observations, F_{obs}. Usually, several cycles of refinement are necessary till the residual differences between F_{obs} and F_{calc} are small. The radius of convergence defines the maximum distance between the position of the atom to be

corrected and its true position. In least squares refinement, $\Sigma_{hkl}(|F_{obs}| - |F_{calc}|)^2$ is minimised, or more properly, the function to be minimised is:

$$\sum_{hkl} w(h\,k\,l)\,(|F_{obs}(h\,k\,l)| - |F_{calc}(h\,k\,l)|)^2$$

the summation is over all crystallographically independent reflections and w is the weight given to an observation.

The progress of refinement can be monitored with the R-factor, which gives an estimation of the errors in a data set:

$$R_{factor} = \frac{\sum_{hkl} ||F_o| - |F_c||}{\sum_{hkl} |F_o|}$$

It represents the summation of the amplitudes of the absolute differences between observed, F_o, and calculated, F_c, structure factors over the sum of the observed amplitudes. It is advisable not to rely entirely on the value of the R-factor as an indication of the success of refinement. Usually, a cross-validation statistic is used, the R-free factor, which is calculated for a separate small portion of the data (5-10%) that is excluded from the refinement process.

2.2.4.4.1 Model refinement of the L226Q HA in complex with LSTa and LSTc

Refinement was carried out using the program REFMAC (Murshudov et al., 1997). This program can be used to perform restrained or unrestrained refinement, in order to minimise differences between F_{obs} and F_{calc}, satisfying a maximum likelihood residual. At the end of

each refinement cycle, a MTZ file is produced, containing weighted coefficients for the calculation of $2F_o$-F_c and F_o-F_c electron density maps (see, below, *2.2.4.5*).

In the present case, REFMAC was also used to create a dictionary entry for the receptor analogue by transferring the corresponding coordinates from a model molecule into that of the target molecule. Structure-based sequence alignments of ligand-less L226Q HA and X-31 HA in complex with LSTa or LSTc, or ligand-less L226Q HA and H3 avian HA in complex with LSTa, were carried out with the program LSQMAN, included in the Uppsala Software Factory (USF) suite of programs. The HA_1 domain comprising residues 59-269 (H3 numbering), which include the HA RBS, was selected for the alignments. The coordinates of the receptor analogues were positioned by these alignments to be transferred to the RBS of the L226Q HA model molecule obtained after running AMoRe (see *2.2.4.2.1.1*). The program LSQMAN was also used in the determination of transformation matrices for the L226Q HA monomers (see *4.5*) and *r.m.s.d* values (see *1.6.6* and *4.7.1.1*). For the calculation of the *r.m.s.d.* on all C_α positions of the 220 loop in the pairwise comparisons L226Q HA/X-31 HA, L226Q HA/H3 avian HA and H3 avian HA/X-31 HA (see *1.6.6*), residues HA_1 224-228 in the monomers of each structure (the coordinates for uncomplexed L226Q HA, X-31 HA and H3 avian HA were obtained from RCSB/PDB, accession codes 3hmg, 1hgf and 1mql, respectively) were used in the alignment and an average value is indicated. In the calculation of the *r.m.s.d.* on all C_α positions in the comparison of RBS1 and RBS2 of L226Q HA in complex with LSTa (see *4.7.1.1*), residues HA_1 59-269 were considered in the alignment.

Several cycles of refinement were carried out, checking the values of working and free R-factors for refinement improvement. Further adjustments were done manually using the molecular graphics program O (Jones et al., 1991) (see, below, *2.2.4.5*), adjusting the

torsion angles of the ligand bonds. For this manual modelling, torsion libraries were generated for LSTa and LSTc by submitting the corresponding *pdb* files to the PRODRG2 server (Schuettelkopf and van Aalten, 2004).

2.2.4.5 Calculation of electron density maps

Electron density maps can be calculated from F_{obs} and α_{calc}. Starting phases can be calculated from a search model, once the molecular replacement parameters are known. The electron density maps calculated using these phases showed if, and how, the receptor analogues LSTa and LSTc are bound on the RBS of L226Q HA. Because experimental amplitudes were used, the maps contain some information about the molecule under study. But as phases tend to dominate in the calculations of electron density maps, these are biased by the model structure. Therefore, the coordinates of the receptor analogue and those of water molecules were not included in the search model for molecular replacement (see *2.2.4.2.1.1*) and in the calculation of omit maps (see, below, *2.2.4.5.3*). This meant that the ligand density had to be interpreted specifically.

Fourier maps were calculated using the program FFT and extended to cover a whole molecule with EXTEND (CCP4 suite). The program MAPMAN (Kleywegt, 1996) was used to convert the calculated maps from the CCP4 format to DSN6, which can be used for display on a silicon Graphics workstation using the program O (Jones et al., 1991).

2.2.4.5.1 F_o-F_c, or difference, Maps

The difference Fourier, or F_{obs}-F_{calc}, maps are obtained from α_{calc}, but have peaks of density not accounted for in the model used to to calculate F_{calc}. They are quite useful to study the binding of ligands to proteins and check the location of water molecules by

analysing the changes introduced in the electron density. However, high *B* values and limited resolution weaken the resolving power of F_o-F_c maps (McRee, 1999).

2.2.4.5.2 $2F_o$-F_c Maps

The $2F_{obs}$-F_{calc} maps are the sum of the electron density obtained from the model and that of a F_{obs}-F_{calc} map. These maps, besides including the electron density corresponding to the protein of the model structure, also enhance the visualization of atoms not included in the phase calculation, like solvent and ligand molecules. Some particular $2F_{obs}$-F_{calc} maps, such as $2mF_o$-DF_c, for which REFMAC calculates coefficients after each refinement cycle, are very useful for fitting. In the study described here, these maps were used for the modelling of LSTa and LSTc bound to the L226Q HA, which allowed the study of the interatomic interactions between protein and receptor analogue (McRee, 1999).

2.2.4.5.3 Omit maps

Omit maps are difference Fourier maps which are calculated by removing from the model specific regions that can then be analysed, in theory, without phase bias from the model. These maps allowed checking the presence of electron density specific for the receptor analogues in complex with the L226Q HA. F_o-F_c and $2F_o$-F_c omit maps were generated for both the complexes L226Q HA-LSTa and L226Q HA-LSTc, by deleting from the model the coordinates of LSTa and LSTc, respectively.

3. Purification and Crystallisation Trials of Full-Length X-31 HA, HA_2 and BHA_2

The aim of this study was to determine the crystal structure of full-length HA in the native conformation, full-length HA_2 in the fusion pH conformation and BHA_2 in the fusion pH conformation (Figure 19) derived from X-31 influenza virus, to get new indications for the membrane fusion mechanism. Throughout the text in this chapter HA, HA_2 and BHA_2 refer to the full-length proteins in the just described conformations. HA contains the transmembrane anchor and palmitoylated HA_2 cytoplasmic tail exposed, HA_2 both the fusion peptide and transmembrane anchor/cytoplasmic tail, and BHA_2 the fusion peptide. Therefore, all the molecules under study have exposed hydrophobic regions prone to non-specific aggregation. These aggregates are usually seen as rosettes upon EM analysis (Ruigrok et al., 1986a). To keep the proteins soluble, a detergent was used throughout the whole purification procedure and also in the crystallisation trials.

The virus ensures a source of large quantities of protein necessary for crystallisation. Also, the presence of oligosaccharide chains on these virus-derived proteins may be important for the formation of inter-trimer contacts, which could not be formed by *E. coli*-expressed molecules (Skehel and Wiley, 2000). HA_2 and BHA_2 were obtained from HA and BHA, respectively, by digestion with serine protease trypsin, which cleaves HA or BHA, in the fusion pH conformation, after HA_1 residues 27 and 224. A disulphide bond links the resulting $HA_1^{28\text{-}224}$ and $HA_1^{225\text{-}328}$ polypeptides. Therefore, HA_1 residues 28-328 were cleaved-off by trypsin treatment of HA and BHA and the remaining HA_1 residues 1-27 were kept on the HA_2 and BHA_2 molecules. The $TBHA_2$ crystal structure revealed at the membrane-distal tip of the molecule the presence of a hydrophobic core, which presumably stabilises the fusion pH conformation, involving some of those HA_1 residues

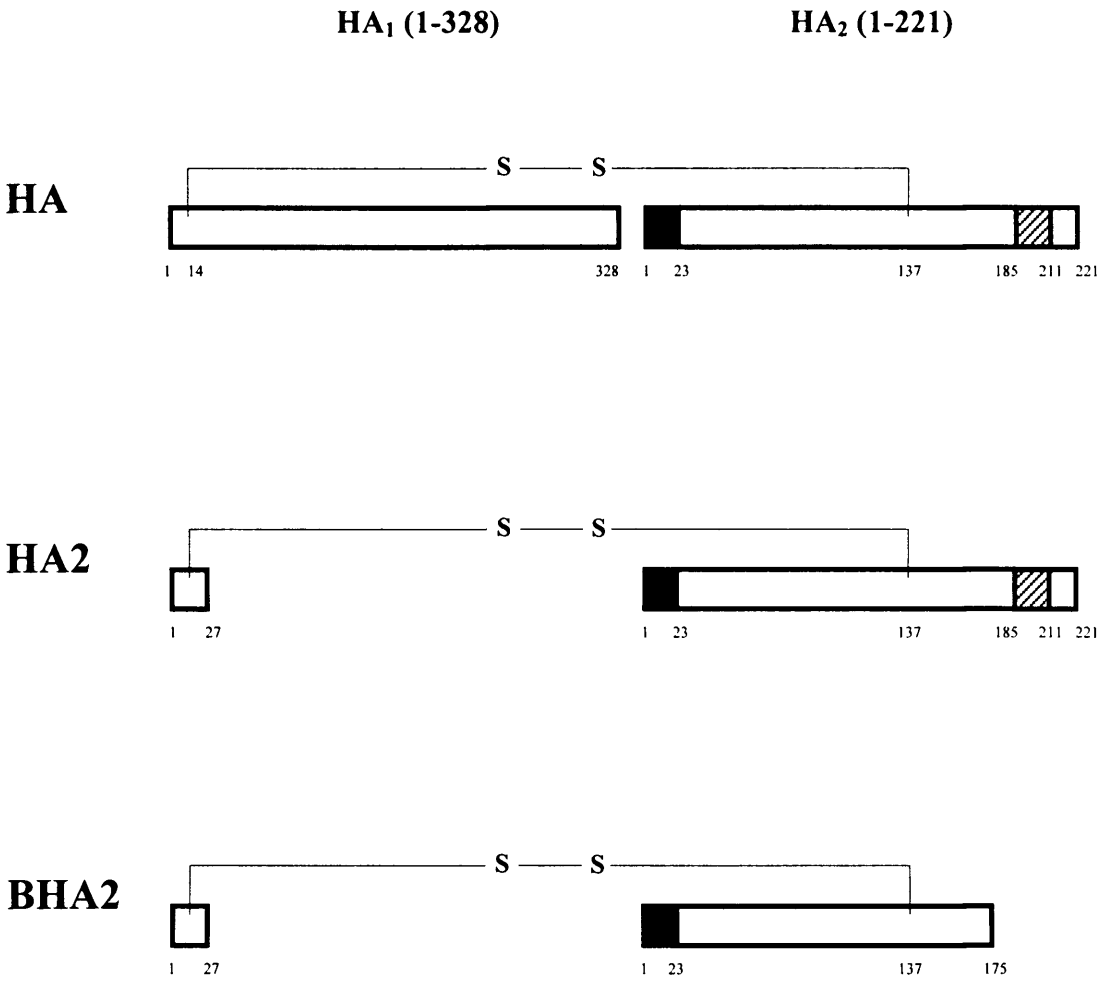


Figure 19. The X-31 HA forms under study

The schematic diagram represents the full-length HA, full-length HA_2 and full-length BHA_2 isolated from X-31 virus and purified as described in this chapter. The black-filled boxes correspond to the fusion peptide and the stripped ones to the transmembrane anchor. The disulfide bond connecting the HA_1 and HA_2 subunits is indicated (see Figure 15).

(Bullough et al., 1994). Therefore, their presence on the studied molecules may be important for the crystallisation process. The absence of transmembrane anchors and lipid-harbouring tails on BHA_2 could render the molecule less hydrophobic, turning its crystallisation easier than that of HA_2. Nonetheless, the structure of BHA_2 could reveal the three-dimensional arrangement of the fusion peptides on the molecule in the fusion pH conformation.

The different steps of the general protocol prepared for the purification of these proteins are described below. Alternative procedures (*3.1.5*, *3.2.2* and *3.2.3*) correspond to unsuccessful or uncompleted experiments. Crystallisation trials were set up for HA, HA_2, and BHA_2 to get hints about suitable conditions for their crystallisation.

3.1 Isolation and Purification of HA and BHA

3.1.1 Isolation of HA from the viral envelope

The full-length HA was solubilised from the viral envelope using the mild nonionic polyoxyethylene detergent $C_{12}E_{10}$ (see *2.1.6*). The solution obtained, after removing the viral cores by ultracentrifugation, contains practically all the viral HA, which has the exposed hydrophobic regions coated by detergent, and also viral membrane lipid – detergent micelles and an excess of pure detergent micelles. Minor apparent low MW contaminations were observed upon SDS-PAGE (Figure 20A). The contaminant polypeptides are most probably breakdown products of the viral membrane proteins and/or membrane remnants from cells of the allantoic fluid where the virus was grown.

3.1.2 Isolation of BHA from the viral envelope

BHA was removed from the influenza virus by cleavage with the cysteine protease bromelain and then separated from the digested virus particles by ultracentrifugation (see

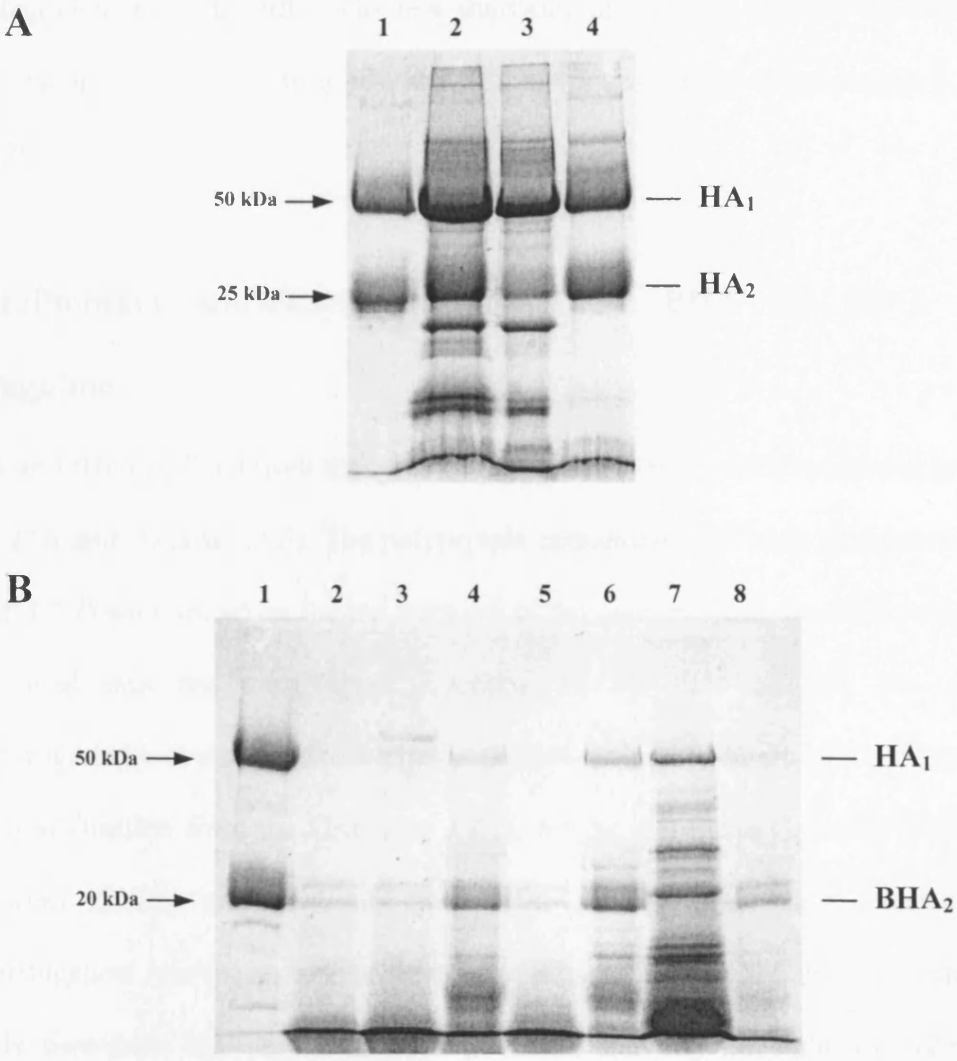


Figure 20. SDS-PAGE of HA and BHA isolated from the influenza virus envelope

(A) Solubilisation of HA from X-31 influenza virus envelope in 2 % $C_{12}E_{10}$. SDS-PAGE of HA under reduced conditions (0.2% 2-ME) on 12 % gels containing 4 M urea, 0.1 % SDS. Lanes: *1*. Marker of pure HA. *2*. Whole virus before incubation with $C_{12}E_{10}$. *3*. Pellet of final solubilisation mixture. *4*. Supernatant of final solubilisation mixture. The viral core ribonucleoprotein co-migrates with HA_1. Therefore, a strong band at that level was obtained in the pellet of the final solubilisation mixture.

(B) Isolation of BHA by bromelain cleavage of X-31 influenza virus. SDS-PAGE of BHA under reduced conditions (0.2% 2-ME) on 12 % gels containing 4 M urea, 0.1 % SDS. Lanes: *1*. Marker of pure BHA. *2*. Initial mixture of 1st digestion. *3*. Final mixture of 1st digestion. *4*. Supernatant of the final mixture of 1st digestion. *5*. Initial mixture of 2nd digestion. *6*. Supernatant of the final mixture of the 2nd digestion. *7*. Pellet of the final mixture of 3rd digestion. *8*. Supernatant of the final mixture of 3rd digestion.

2.1.7). The resulting SNs contain BHA and bromelain, which shows low MW smears due to auto-degradation (Figure 20B). The first short step of digestion enabled the removal of the majority of the contaminating NA, which is more susceptible to bromelain digestion (see *3.1.3*).

3.1.3 Preliminary purification of HA and BHA by SDG ultra-centrifugation

Both HA and BHA isolated from the virus were purified in linear 5-25 % sucrose gradients (Figures 21A and B) (see *2.1.8*). The polypeptide contaminants of both viral proteins (see *3.1.1* and *3.1.2*) were found on the top fractions of the gradients and, for BHA, also 2-ME was removed into the most upper fractions. In the case of HA, the gradient ultracentrifugation was also used to exchange the non-dialysable detergent $C_{12}E_{10}$, used for the HA solubilisation from the virus (see *3.1.1*), for the dialysable C_{12}DAO, which only differs from $C_{12}E_{10}$ on its head group. The detergent exchange during SDG ultracentrifugation was necessary because, the subsequent step of dialysis required a dialysable detergent. HA- and BHA-rich fractions, showing no signs of polypeptide contaminants upon gel analysis (fractions 4, 5 and 6 in Figures 21A and 21B), were selected from the collected gradient fractions. No BHA fractions were selected from the gradients corresponding to the SN from the first bromelain digestion, as all fractions had high or very high levels of NA (data not shown). Also, no HA fractions below fraction 4 (see Figure 21B) were selected as they could contain HA aggregates. The selected HA and BHA gradient fractions were then pooled and dialysed to remove the sucrose (see *2.1.9*). Figure 21C shows the polyacrylamide gel analysis of the purified proteins, both in the dialysis buffer: HA is in 0.1 % C_{12}DAO, 10 mM Tris pH 8.0, and BHA in 10 mM Tris pH 8.0.

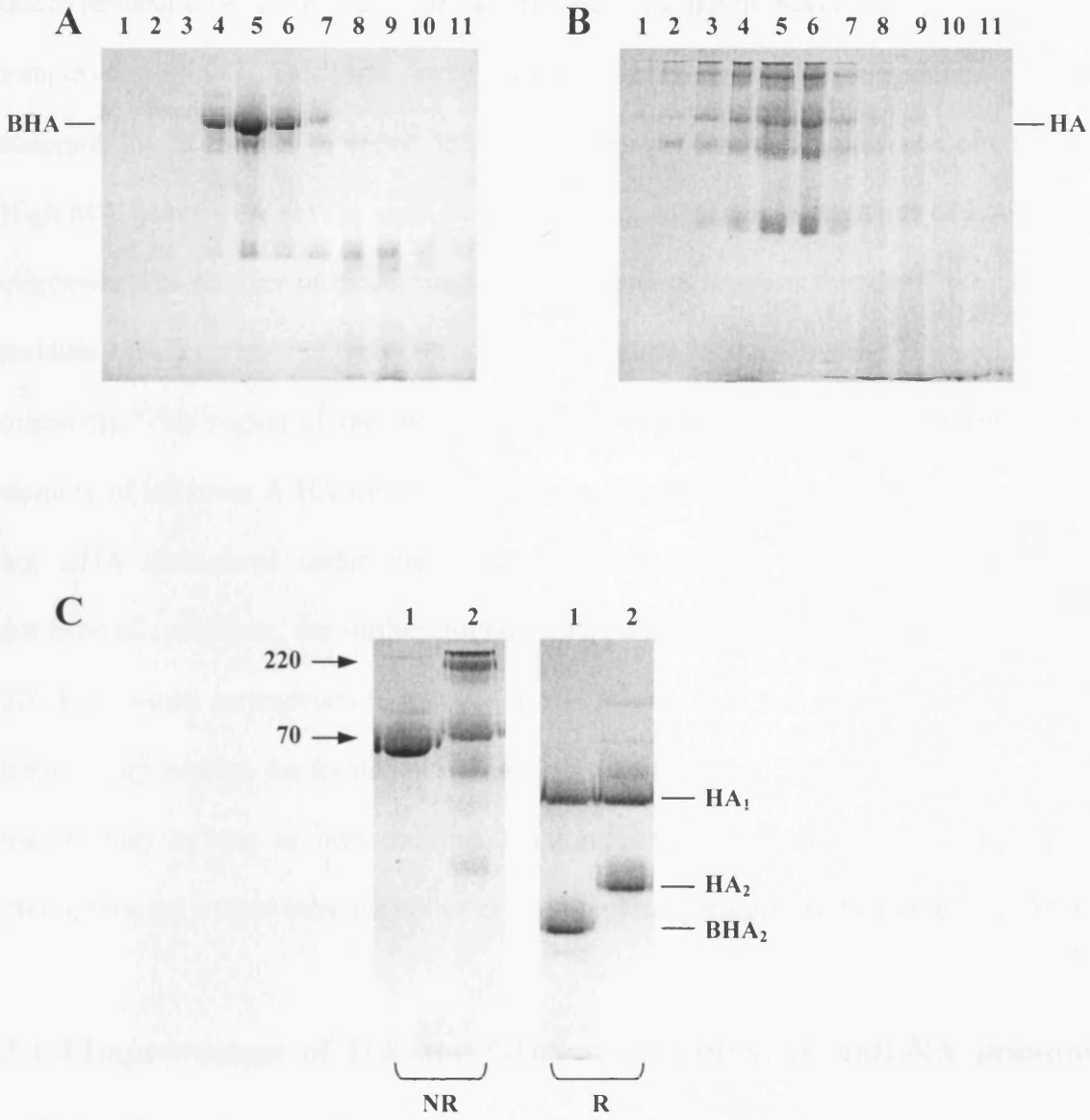


Figure 21. SDS-PAGE of HA and BHA purified by SDG ultracentrifugation

Purification of BHA and HA by SDG ultracentrifugation on linear 5-25 % sucrose gradients. SDS-PAGE on 12 % gels containing 4 M urea, 0.1 % SDS. BHA (*A*) or HA (*B*) gradient fractions. Lanes: *1-11*. Gradient fractions 1-11, respectively. Fractions were numbered in crescent order from bottom to top of gradients. (*C*) Purified HA and BHA after dialysis and concentration by centrifugation (see *2.1.10*). Lanes: *1*. BHA. *2*. HA. R: reduced conditions (0.2% 2-ME). NR: nonreduced conditions.

Faint bands for HA_1 and HA_2 were frequently seen in nonreduced conditions, suggesting that a residual level of reduction of the HA_1-HA_2 disulphide bound has occurred during sample preparation. This was more evident for HA samples, suggesting a higher susceptibility of the HA molecule to reduction, maybe due to the presence of detergent. High MW bands were always seen for HA samples, suggesting a tendency of HA to form oligomers. The absence of those bands in BHA samples suggests that the C-terminal HA_2 residues 176-221, removed by bromelain, are important for the formation or stability of the oligomers. This region of the molecule has previously shown to be important for the stability of influenza A HA trimers. HA trimers remained intact upon incubation in SDS, but BHA dissociated under the same conditions (Doms and Helenius, 1986). In nonreduced conditions, the further migrating HA oligomer band has an apparent MW of 220 kDa, which corresponds to that of the HA trimer. This fact suggests that stable HA trimers may mediate the formation of higher order oligomers. The higher stability of HA trimers may be due to intra-transmembrane anchor interactions (Tatulian and Tamm, 2000), forming a triple-stranded coiled coil as previously suggested (Wilson et al., 1981).

3.1.4 Improvement of HA and BHA purification by anti-NA immuno-affinity chromatography

The detection of low concentrations of NA in HA or BHA samples analysed in polyacrylamide gels is hampered by the fact that some NA bands co-migrate with HA, as checked by running the two proteins separately in gel (data not shown). Therefore, a colorimetric assay testing for NA activity (see *2.1.12*) was done to check if there was still NA present in the samples purified by SDG ultracentrifugation (see *3.1.3*). Both post-dialysis HA and BHA solutions showed significant levels of NA activity. Immunoaffinity chromatography was then used to remove the contaminating NA (see *2.1.11*). The

available anti-NA mAb showed to have low affinity for NA. Extensive passage of HA and BHA samples was performed both at RT and 4°C. The dialysis buffers (the protein buffers of the samples used) 0.1 % C_{12}DAO, 10 mM Tris pH 8.0, for HA, and 10 mM Tris pH 8.0, for BHA, were used to pre-equilibrate and wash the columns. Even though the levels of NA activity could be reduced, the activity was not totally eliminated. Also, significant losses of protein sometimes occurred. This could have been caused by protein aggregation inside the column.

3.1.5 Alternative purification of HA and BHA by one-single-step IE chromatography

Alternatively, a simpler procedure for the purification of the isolated HA and BHA was performed using only IE spin columns. In the case of the HA, the SN containing the solubilised protein in 2% $C_{12}E_{10}$, PBS, 0.01% azide was passed through the IE column, after dilution with 2% $C_{12}E_{10}$, 10 mM Tris pH 8.0 to reduce the salt content to a minimum of 30 mM. The IE column was then washed with 2% $C_{12}E_{10}$, 10 mM Tris pH 8.0 (washing buffer), ensuring the removal of any remaining viral membrane lipids, and the HA eluted with 2% $C_{12}E_{10}$, 10 mM Tris pH 8.0, 400 mM NaCl (elution buffer). For BHA, the pooled SNs from the bromelain digestion were directly passed through the IE column. An identical procedure was followed as for HA, using 10 mM Tris pH 8.0 as the washing buffer and 10 mM Tris pH 8.0, 400 mM NaCl as the elution buffer. An extensive washing of the column-bound BHA ensured the complete elimination of 2-ME. Both HA and BHA samples were highly concentrated by eluting them in the minimum possible volume. Overloaded lanes in SDS-PAGE gels showed faint contaminant bands in HA and BHA samples (Figures 22A and B).

Therefore, this seems to be a valid simple procedure for the purification of HA and BHA,

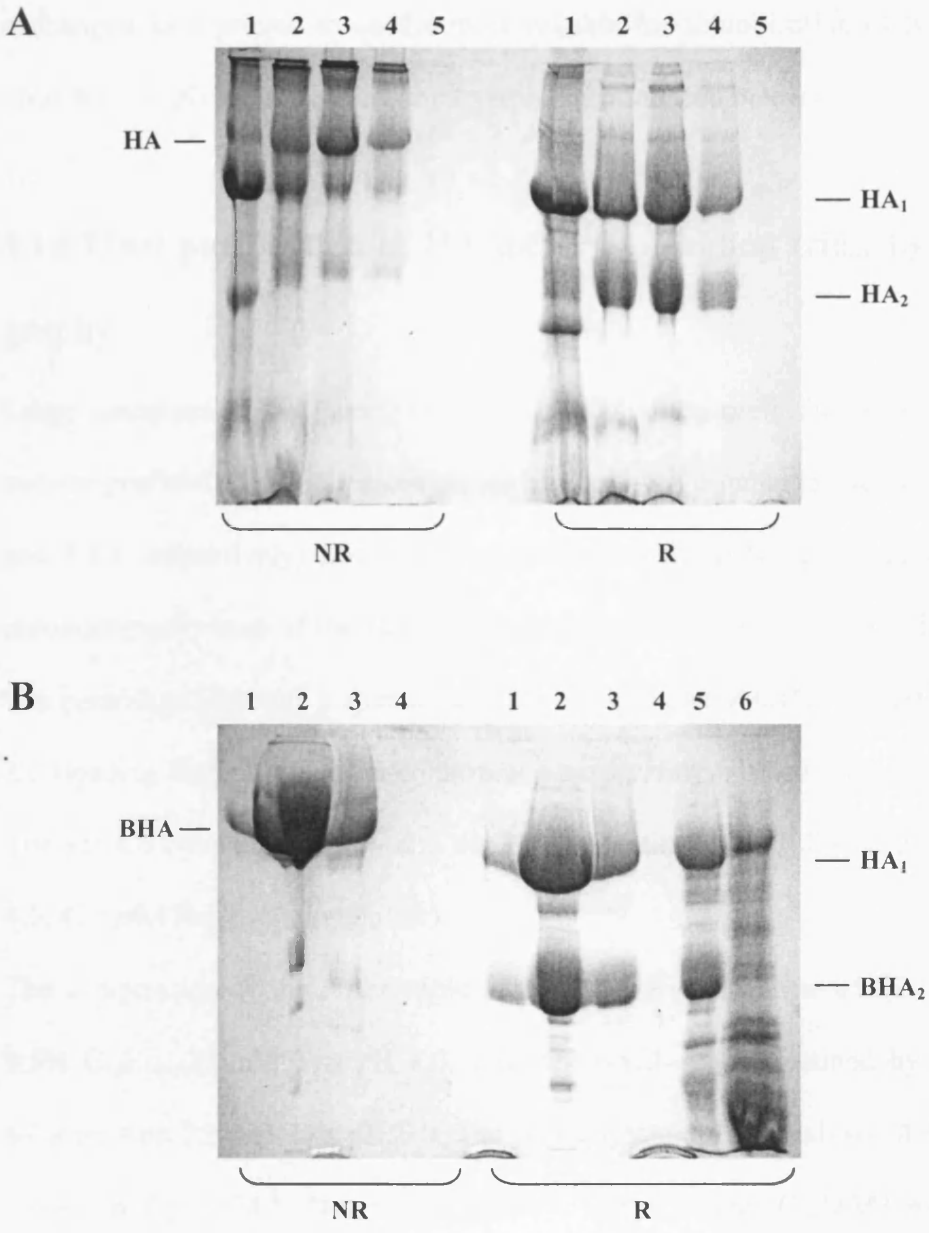


Figure 22. SDS-PAGE of HA and BHA purified by IE chromatography

Purification of HA and BHA by IE chromatography. SDS-PAGE on 12 % gels containing 4 M urea, 0.1 % SDS. Samples subjected to IE, detergent-solubilised HA (*A*) or bromelain-released BHA (*B*), were eluted with 400 mM NaCl. (*A*) Lanes: *1*. Pellet of final solubilisation mixture; *2*. SN of final solubilisation mixture. *3*. 1st elution from the IE column (50 μg, loaded protein); *4*. 2nd elution from the IE column; *5*. Flow-through from IE column. (*B*) Lanes: *1*. Marker of BHA; *2*. 1st elution from IE column (80 μg, loaded protein); *3*. 2nd elution from the IE column; *4*. Flow-through from IE column; *5*. SN of ON digestion; *6*. Pellet of ON digestion. R: reduced conditions (0.2% 2-ME). NR: nonreduced conditions. No bands were obtained in the flow-through, as samples had to be diluted to reduce the ion strength before passage through the column.

but it needs to be further developed. The detergent $C_{12}E_{10}$ in HA samples was not exchanged as it proved to be the most suitable for solubilisation of both HA and BHA upon the low pH incubation for the trypsin digestion (see below).

3.1.6 Final purification of HA for crystallisation trials by IE chromatography

Large quantities of HA were solubilised from virus, preliminarily purified in a linear sucrose gradient, dialysed, passed through an anti-NA column (as described in *3.1.1*, *3.1.3*, and *3.1.4*, respectively) and, finally, passed through an IE spin column. In this last IE chromatography step of the HA purification protocol (see *2.1.13*), firstly, the HA sample was centrifuged through a spin column pre-equilibrated in 0.1% C_{12}DAO, 10 mM Tris pH 8.0 (loading buffer). Then the column was extensively washed with 1.5% $C_{12}E_{10}$, 25 mM Tris pH 8.0 (washing buffer), and the HA was eluted with 1.5% $C_{12}E_{10}$, 25 mM Tris pH 8.0, 450 mM NaCl (elution buffer).

The composition of the HA sample used in the crystallisation trials – 10 mg/ml HA in 0.5% $C_{12}E_{10}$, 25 mM Tris pH 8.0, 150 mM NaCl – was obtained by diluting the elute solution with 25 mM Tris pH 8.0. The polyacrylamide gel analysis of this HA sample is shown in Figure 28A, lane 1 (see below). The detergent C_{12}DAO was exchanged for $C_{12}E_{10}$, which gave indications to be the most suitable for solubilisation of HA_2 and BHA_2 (see *3.2.4*), and also to keep all proteins solubilised in the same detergent for the initial crystallisation trials. The ionic strength was set to 150 mM NaCl to use the same value of the solutions of BHA proteins derived from other influenza virus subtypes under crystallisation in the laboratory [L. F. Haire (Division of Molecular Structure, MRC-NIMR, London), personal communication]. Common crystallisation patterns may be

derived from the comparison of the different proteins, helping the identification of promising crystallisation conditions.

3.2 Isolation and Purification of HA_2 and BHA_2

HA_2 and BHA_2 were obtained by trypsin digestion of low pH-treated HA and BHA in detergent. HA was already in detergent as it contains the hydrophobic transmembrane anchor and palmitate residues. During the required previous low pH treatment, both HA and BHA extrude the fusion peptide. Therefore detergent had also to be present in BHA and BHA_2 samples to keep the proteins soluble. Samples of HA and BHA isolated from virus, initially purified by SDG ultracentrifugation and passed through an anti-NA column, as described above (see *3.1.1*, *3.1.3*, and *3.1.4*, respectively), were used for all procedures, unless otherwise indicated.

3.2.1 Trypsin digestion of HA and BHA in the detergent $C_{12}DAO$

The first test digestions were performed in the detergent $C_{12}DAO$. Several different conditions were tried for the low pH and trypsin incubation steps. The proteins, at 2 mg/ml in 10 mM Tris pH 8.0, were incubated for 10 min at pH 5.0 or 4.8 in the presence of $C_{12}DAO$ at concentrations ranging 0.1-1.1 % and then incubated for 30 min at pH 7.0 with trypsin, at trypsin:HA or BHA w/w ratios of 1:50, 1:40 or 1:20 (see *2.1.14*). Upon gel analysis of the reaction mixtures, the presence of a HA_1 band in all the conditions indicated that the reaction was not complete. Also, a band migrating a little further than that of HA_2 or BHA_2 suggested that the N-terminus was being cleaved at a labile region as previously described (Ruigrok et al., 1988). This over-digestion was more apparent, after incubation at pH 4.8, at $1:20_{w/w}$ trypsin:HA or BHA ratios (Figure 23A). To overcome this problem, new conditions were tried, following the indication that the N-terminus of HA_2 was not

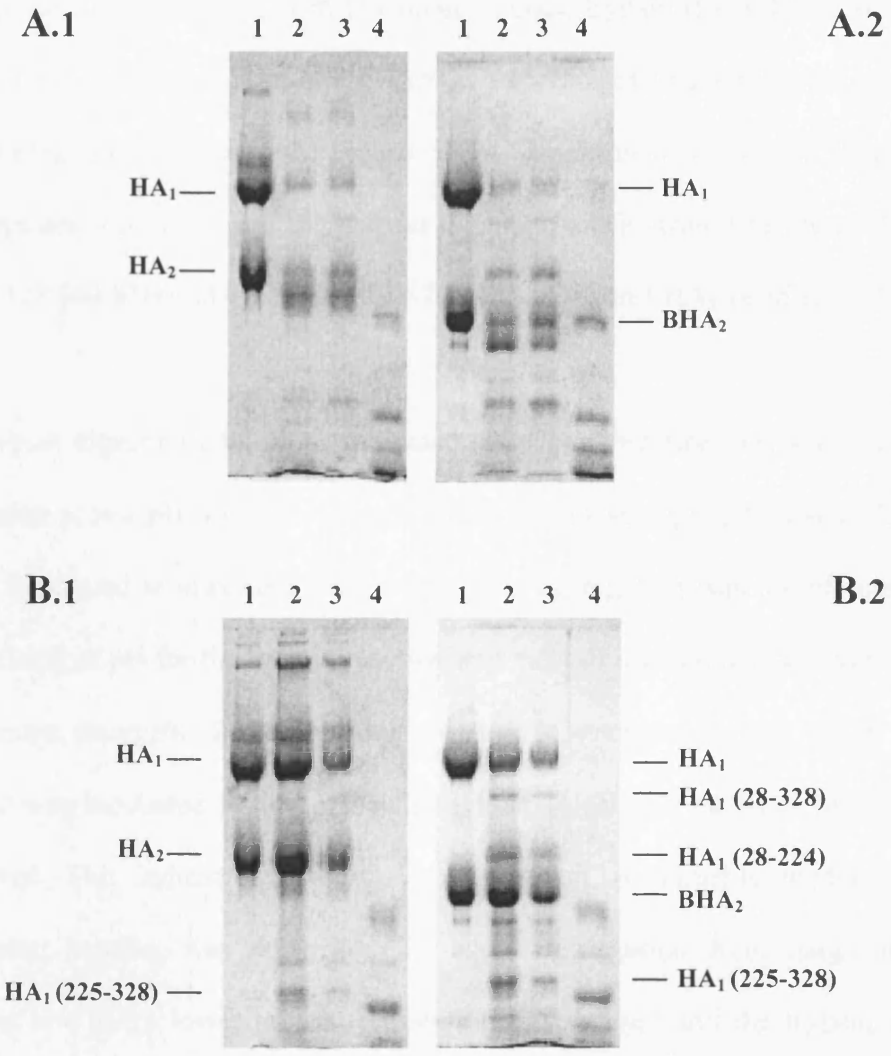


Figure 23. SDS-PAGE of HA and BHA digested with trypsin in C_{12}DAO

(A) Digestion of HA and BHA with trypsin at pH 7.0 in C_{12}DAO. SDS-PAGE on 12 % gels containing 4 M urea, 0.1 % SDS of HA (*A.1*) and BHA (*A.2*), under reduced conditions (0.2% 2-ME), after digestion with trypsin. (*A.1*) Lanes: *1*. HA marker; *2*. Final reaction mixture*; *3*. Final reaction mixture passed by a TI column; *4*. Trypsin marker. (*A.2*) Lanes: *1*. BHA marker; *2*. Final reaction mixture*; *3*. Final reaction mixture passed by a TI column; *4*. Trypsin marker. *: samples were frozen after reaction.

(B) Digestion of HA and BHA with trypsin at pH 6.0 in C_{12}DAO. SDS-PAGE on 12 % gels containing 4 M urea, 0.1 % SDS of HA (*B.1*) and BHA (*B.2*), under reduced conditions (0.2% 2-ME), after digestion with trypsin. (*B.1*) Lanes: *1*. HA marker; *2*. Final reaction mixture*; *3*. Final reaction mixture passed by a TI column; *4*. Trypsin marker. (*B.2*) Lanes: *1*. BHA marker; *2*. Final reaction mixture*; *3*. Final reaction mixture passed by a TI column; *4*. Trypsin marker. *: samples were frozen after reaction.

cleaved by trypsin at pH 6.0 (Ruigrok et al., 1988). HA and BHA were incubated for 10 min at pH 5.0 and then incubated with trypsin at $1:100_{w/w}$ trypsin:HA or BHA for one hour at pH 6.0. Under these conditions, even though the reaction was far from complete, no significantly clear band indicative of HA_2 or BHA_2 digestion was obtained (Figure 23B). The HA_1 peptides resulting from the trypsin digestion are indicated in Figure 23B: HA_1 residues 28-328 (40 kDa), HA_1 residues 28-224 (25 kDa) and HA_1 residues 225-328 (15 kDa).

In all the trypsin digestion conditions indicated above, aggregation was visually detected upon incubation at low pH, for both HA and BHA. Extensive aggregation was obtained in the samples incubated at lower pH (pH = 4.8). This aggregation was not affected by the subsequent rising of pH for the trypsin reaction and most of it could not be reversed by the addition of more detergent. To check if this was due to detergent coming out of solution, 1% C_{12}DAO was incubated at low pH. After 1 hour at pH 4.7, no signs of aggregation were observed. This indicated that protein aggregation, presumably mediated by the exposed fusion peptide, was occurring. To avoid aggregation from happening upon incubation at low pH, a lower protein concentration was used and the trypsin digestion carried out in the conditions shown not to induce N-terminal HA_2 cleavage, as described above. Protein samples at 0.3 mg/ml in 0.1% C_{12}DAO incubated for 10 min at pH 5.0 showed occasional small-scale aggregation.

3.2.2 Purification of HA_2 by anti-HA_1 immunoaffinity chromatography

To remove the trypsin-cleaved HA_1 40 kDa fragment and the uncleaved full-length HA, final reaction mixtures of the trypsin digestion of HA in C_{12}DAO showing no signs of aggregation were passed through a mAb HC73 affinity column. HC73 binds to both the low pH and native conformation of HA, at HA_1 residues 134 and 145. No protein could be

detected in the void volume and HA_2, free of the HA_1 40 kDa fragment and full-length HA, was recovered in the subsequent fractions collected by passing 10 mM Tris pH 8.0, 0.1% C_{12}DAO through the column (Figure 24). The fact that HA_2 was only eluted after the void volume could be a consequence of the protein having to pass through the pores of sepharose beads. The positioning of a protein between the void and the included volumes of a gel filtration column is a criterion for protein solubility (Boehringer Mannheim, 1990). But it could also reflect the aggregation of HA_2 inside the column. In fact, aggregation was observed in some of the eluted samples, even though the protein concentration was as low as 0.1 mg/ml.

3.2.3 Purification of HA_2 and BHA_2 by SDG ultracentrifugation

The purification of trypsin-isolated HA_2 and BHA_2 in C_{12}DAO was also tried by ultracentrifugation in linear sucrose gradients (see *2.1.8*). The final trypsin digestion reaction mixtures showing no signs of aggregation (see *3.2.1*) were ultracentrifuged in linear 5-15% sucrose gradients with a 60% sucrose cushion. The uncleaved full-length HA or BHA were expected to be found at the 15/60% sucrose interface. The polyacrylamide gel analysis of the gradient fractions showed that the different polypeptides did not form sharply defined bands, leading to their overlapping (data not shown). This made the selection of pure HA_2 or BHA_2 bands not possible.

3.2.4 Trypsin digestion of HA and BHA in the detergent $C_{12}E_{10}$

The detergent C_{12}DAO of HA samples in 0.1 % C_{12}DAO, 10 mM Tris pH 8.0 was exchanged for $C_{12}E_{10}$ by IE chromatography (see *2.1.13*). Each HA sample in C_{12}DAO was passed through a pre-equilibrated IE spin column, extensively washed with 2% $C_{12}E_{10}$, 10 mM Tris pH 8.0 (washing buffer) and eluted with 2% $C_{12}E_{10}$, 10 mM Tris pH 8.0, 400

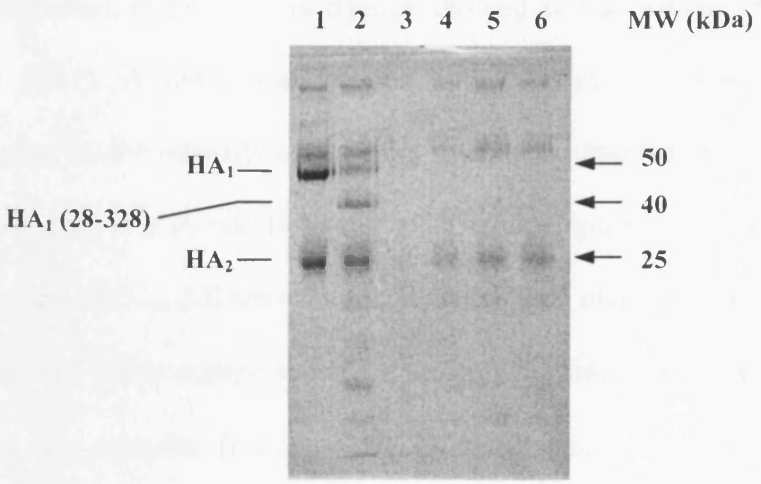


Figure 24. SDS-PAGE of HA_2 purified by immunoaffinity chromatography

Purification of HA_2 by immunoaffinity chromatography. SDS-PAGE, under reduced conditions (0.2% 2-ME), on 12 % gels containing 4 M urea, 0.1 % SDS of trypsin-digested HA passed through an anti-HA_1 column. Lanes: *1*. Marker of pure HA; *2*. Trypsin digestion final mixture; *3*. Void volume from anti-HA_1 column; *4-6*. 1^{st}-3^{rd} fraction of recovered HA_2, respectively.

mM NaCl.

Samples of HA or BHA at up to 1.0 mg/ml in 1 or 2 % $C_{12}E_{10}$ showed no signs of aggregation even after ON incubation at pH 5.0. The conditions for the digestion of HA and BHA with trypsin in $C_{12}E_{10}$ were then extensively screened, aiming at a complete cleavage, which would simplify the purification of HA_2 and BHA_2. The trypsin digestions were always performed at pH 6.0, as trypsin showed not to cleave HA_2 under those conditions (see *3.2.1*). A time-course screen showed that a maximum of digestion extension, evaluated by the intensity of the HA_1 band, was obtained after about 2 hours of incubation with trypsin at a trypsin:HA ratio of $1:50_{w/w}$ (Figure 25A). An increase of the trypsin:HA ratio up to $1:5_{w/w}$ did not affect the extension of digestion (data not shown). A pH screen of the low pH treatment showed that after incubation of HA at pH $\leq$ 4.8 the trypsin digestion was complete (Figure 25B.1). A smaller-scale screen for BHA showed that after incubation at pH $\leq$ 4.5 the digestion was also complete (Figure 25B.2). These results indicated that an incubation at a pH $\leq$ 4.8 was necessary to induce a conformational change in all the molecules in solution. The HA_1 peptides obtained from the trypsin digestion are indicated in Figure 25: HA_1 residues 28-328 (40 kDa), HA_1 residues 28-224 (25 kDa) and HA_1 residues 225-328 (15 kDa).

Upon incubation at lower pH values, the HA_1 40 kDa fragment is practically no longer detected, suggesting that it is susceptible to further trypsin digestion under those conditions. The split of the HA_1 40 kDa band into the 25 kDa and 15 kDa bands, by reduction of the connecting disulphide bond, is obvious for the BHA-derived HA_1 but not for the HA-derived. This was due to co-migration of the HA_1 25 kDa fragment with the HA_2 monomer in trypsin-digested HA samples. Both HA_2 and BHA_2 showed to be stable at pH values as low as 3.4 and 4.0, respectively, and no aggregation was observed in any of the samples.

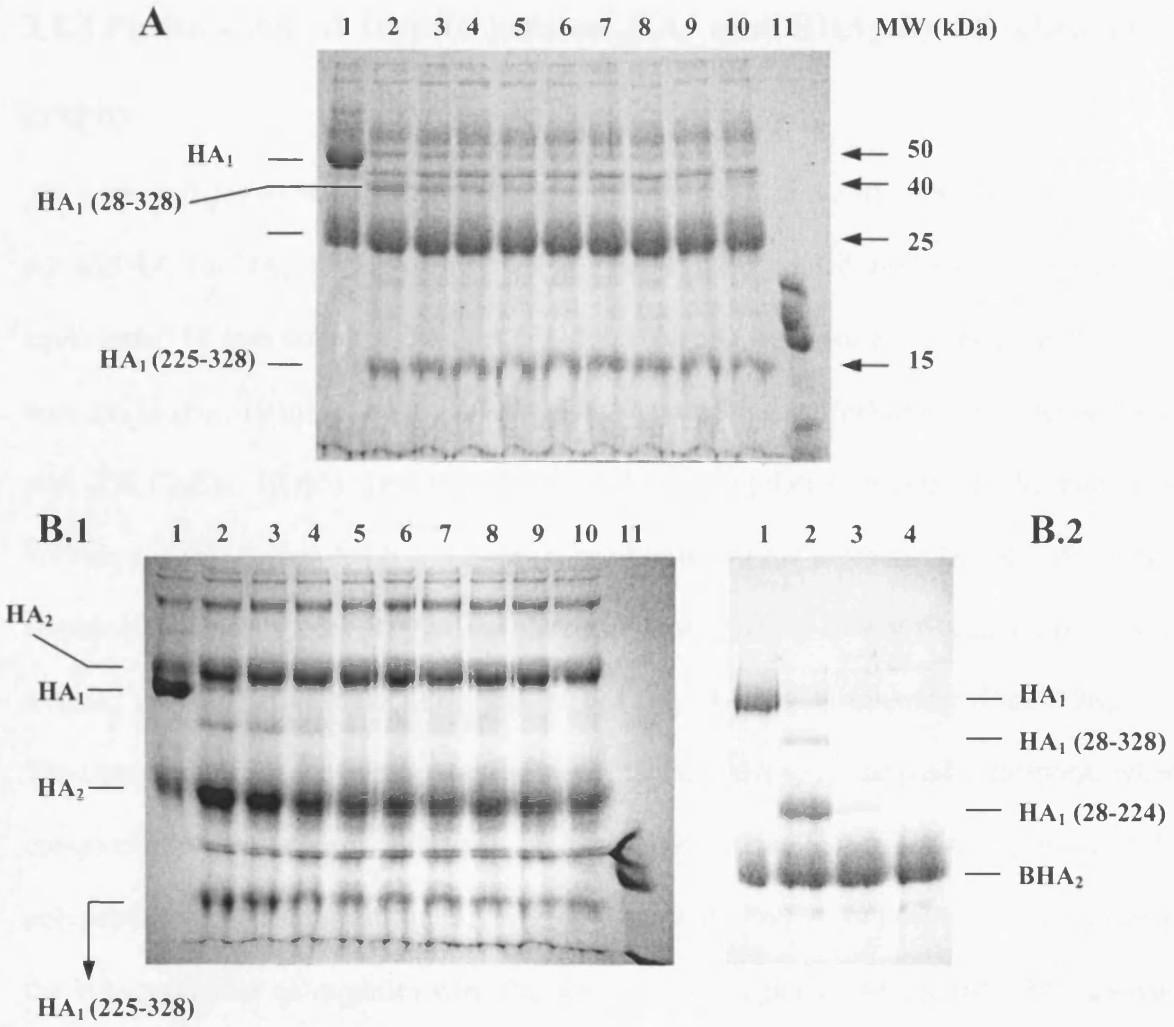


Figure 25. SDS-PAGE of HA and BHA from the screening of trypsin digestion conditions in $C_{12}E_{10}$

(*A*) Time course screen of the conditions for the digestion of HA with trypsin in $C_{12}E_{10}$. SDS-PAGE of HA, under reduced conditions (0.2% 2-ME), on 12 % gels containing 4 M urea, 0.1 % SDS. SNs of solubilised HA at 1 mg/ml were incubated at pH 5.0 for 40 min and then incubated with trypsin at pH 6.0. Fractions were taken at successive times of ½ hour and immediately treated with trypsin inhibitor. Lanes: *1.* HA solubilisation SN after incubation at low pH; *2-10.* Trypsin digestion mixture at successive times; *11.* Mixture $1{:}1_{w/w}$ of trypsin and trypsin inhibitor. The sharp band seen slightly above that of HA_1 residues 28-328 may correspond to an oligomer of HA_1 residues 225-328 obtained is this particular occasion.

(*B*) pH screen of the low pH treatment of HA and BHA previous to trypsin digestion in $C_{12}E_{10}$. SDS-PAGE on 12 % gels containing 4 M urea, 0.1 % SDS of HA (*B.1*) and BHA (*B.2*) samples, under reduced conditions (0.2% 2-ME). HA and BHA samples at 1 mg/ml in 2% $C_{12}E_{10}$, 400 mM NaCl were incubated for 95 min at different pH values and then incubated for 3 ½ hours at pH 6.0 with trypsin at $1{:}50_{w/w}$. (*B.1*) Lanes: *1.* Marker of pure HA; *2-10.* Trypsin digestion final mixture after incubation at different pH values: 5.0, 4.8, 4.6, 4.4, 4.2, 4.0, 3.8, 3.6 and 3.4, respectively; *11.* Mixture $1{:}1_{w/w}$ of trypsin and trypsin inhibitor. The band seen above that of HA_1 as well as the ones at the top of the gel correspond to oligomers of HA_2 (see *3.2.6*). (*B.2*) Lanes: *1.* Marker of BHA; *2-4.* Trypsin digestion final mixture after incubation at different pH values: 5.0, 4.5 and 4.0, respectively.

3.2.5 Purification of trypsin-isolated HA_2 and BHA_2 by IE chromatography

The trypsin-digested samples from the low pH screens (see *3.2.4*) after incubation at pH 4.2 and 4.4, for HA, or pH 4.0 and 4.5, for BHA, were pooled and passed through pre-equilibrated IE spin columns (see *2.1.13*). The columns were pre-equilibrated and washed with 2% $C_{12}E_{10}$, 10 mM Tris pH 6.0 (loading and washing buffer) and the proteins eluted with 2% $C_{12}E_{10}$, 10 mM Tris pH 6.0, 500 mM NaCl (elution buffer) in the minimum volume possible, which led to a 5 times concentration of the proteins. The eluted samples contained practically pure HA_2 or BHA_2 with minor apparent low MW contaminations of trypsin, TI and putative break down products of the HA_1 40 kDa fragment (Figure 26).

The observed difference in migration of both HA_2 and BHA_2 in reduced conditions, when compared to nonreduced conditions, reflects the release of the HA_1 residues 1-27 polypeptide by reduction of the $HA_1 - HA_2$ disulphide bound. In nonreduced conditions, the HA_2 monomer co-migrates with the HA_1 40 kDa fragment. As for HA, HA_2 showed bands of higher MW than that of the monomer, suggesting oligomer formation. As obtained for BHA, this was not observed for BHA_2, further suggesting that the region, or part of it, cleaved off by bromelain is important for the formation or stability of the oligomers. The analysis of band migration does not clearly suggest the oligomerisation order of the formed oligomers. The comparison of the lanes for HA and HA_2, in reduced conditions, in Figure 26A suggests the oligomers seen for the HA sample may in fact have been formed by HA_2, which was obtained from HA upon sample preparation for gel analysis.

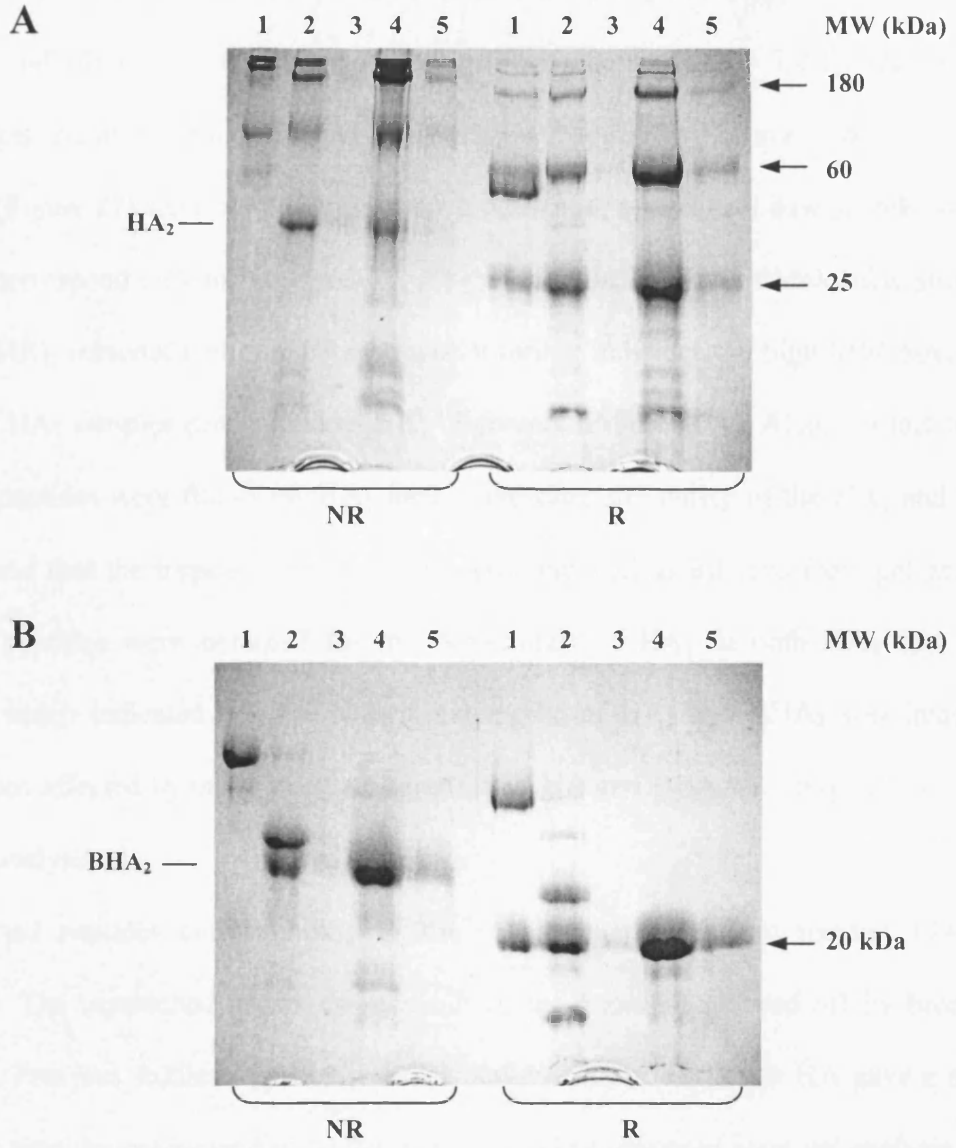


Figure 26. SDS-PAGE of HA_2 and BHA_2 purified by IE chromatography

Purification of HA_2 and BHA_2 by IE chromatography. SDS-PAGE on 12 % gels containing 4 M urea, 0.1 % SDS of HA_2 (*A*) and BHA_2 (*B*). Lanes: *1*. Marker of pure HA or BHA; *2*. Trypsin digestion final mixture; *3*. Flow-through; *4*. 1st elution; *5*. 2nd elution. R: reduced conditions (0.2% 2-ME). NR: nonreduced conditions. In (A), lane 4/R the 60kDa band corresponds to the mass spectrum shown in Figure 27.

(No bands were observed in gel for the flow-through, as samples had to be diluted to reduce ionic strength before passing by the column)

3.2.6 Analysis of HA_2 and BHA_2 by MALDI spectrometry

The HA_2 and BHA_2 samples eluted from the IE spin columns (see *3.2.5*) were run in a different gel and all the bands analysed by MALDI spectrometry (see *2.1.16*). The spectra obtained (Figure 27) were very similar to each other and, after a data base search, all were found to correspond only to HA_2 (see *2.1.16*) [S. Howell (Division of Molecular Structure, MRC-NIMR), personal communication], which further indicates the high MW bands seen in gel for HA_2 samples correspond to HA_2 oligomers (Figure 26A). Also, the fact that no matched peptides were found for HA_1 further indicates the purity of the HA_2 and BHA_2 samples and that the trypsin digestion had been complete, as inferred from gel analysis. Matched peptides were obtained for the N-terminus of HA_2 in both HA_2 and BHA_2 samples, which indicated that the N-terminal region of HA_2 and BHA_2 was intact, not having been affected by the performed digestion of HA and BHA with trypsin, as inferred from gel analysis.

No matched peptides corresponding to the HA_2 sequence beyond residue 174 were identified. The unmatched region corresponds to the fragment cleaved off by bromelain from HA. Previous studies of protease susceptibility of the full-length HA gave a similar result. No peptides corresponding to that region could be observed upon gel analysis [S. A. Wharton (Division of Virology, MRC-NIMR), personal communication]. This may reflect the fact that this region is not, as the exposed ectodomain, under selective pressure for protection against degradative enzymes. Taken together, these results suggest a higher susceptibility to unspecific proteolysis for the HA_2 residues 175-221 stretch. Therefore, direct evidence that the trypsin-isolated HA_2 has an intact C-terminus is not available. But the fact that upon gel analysis in reduced conditions its migration can not be distinguished from that of HA_2 obtained by reduction of control HA in sample preparation (Figure 26A) suggests the protein is full-length intact.

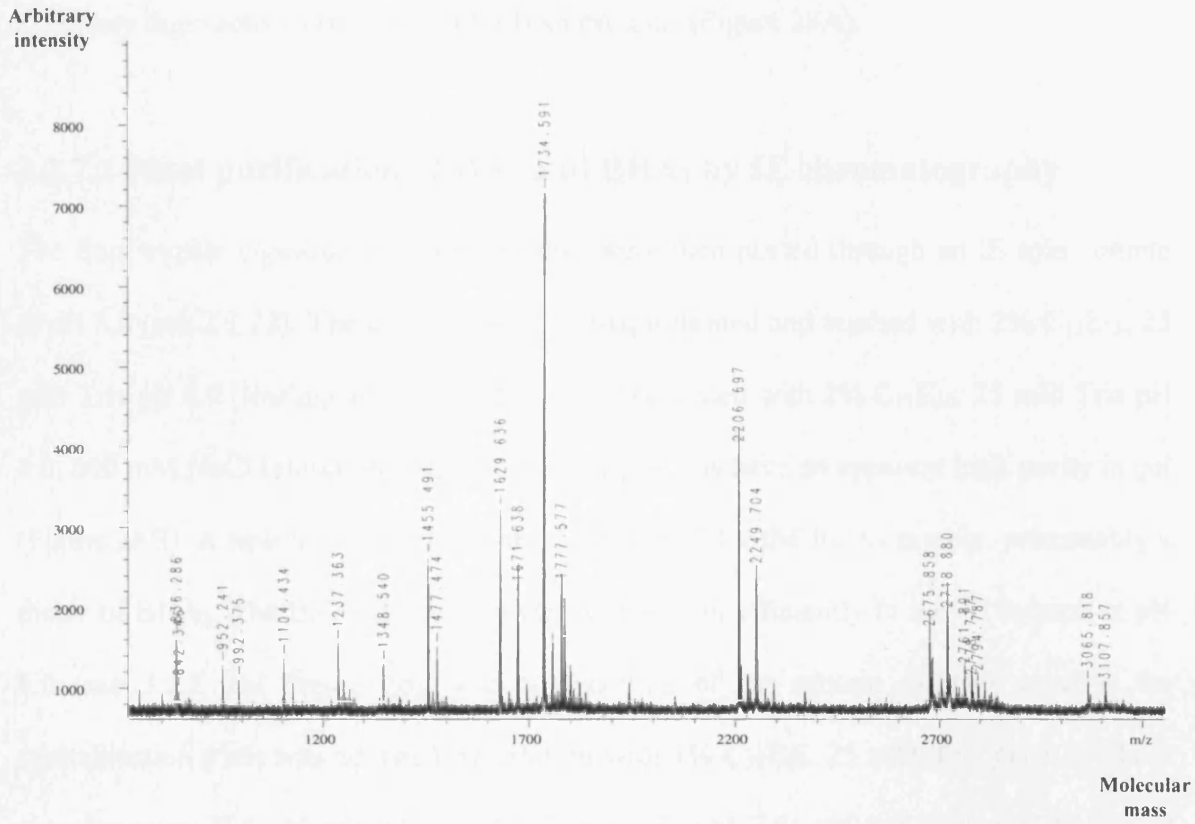


Figure 27. MALDI spectrometry analysis of HA_2

MALDI spectrometry analysis of HA_2 purified by IE chromatography.

A representation of arbitrary intensity as a function of molecular mass is shown.

The peptide mass spectrum of in-gel-digested HA_2 is shown. The gel analyses of the HA_2 and BHA_2 samples processed for MALDI spectrometry are shown in Figure 26. The band whose spectrum is shown here is identified in the previous figure. The spectra obtained for BHA_2 and the other HA_2 bands have minor differences comparing to the one shown.

3.2.7 Preparation of HA_2 and BHA_2 for crystallisation trials

3.2.7.1 Optimised trypsin digestion of HA_2 and BHA_2 in $C_{12}E_{10}$

HA or BHA at 1 mg/ml in 2% $C_{12}E_{10}$ were incubated for 2 ½ hours at pH 4.6 and then incubated with trypsin, $1{:}50_{w/w}$ trypsin:HA or BHA, for 4 hours at pH 6.0 (see *2.1.14*). Complete digestions were obtained for both proteins (Figure 28A).

3.2.7.2 Final purification of HA_2 and BHA_2 by IE chromatography

The final trypsin digestion reaction mixtures were then passed through an IE spin column at pH 8.0 (see *2.1.13*). The columns were pre-equilibrated and washed with 2% $C_{12}E_{10}$, 25 mM Tris pH 8.0 (loading and washing buffer) and eluted with 2% $C_{12}E_{10}$, 25 mM Tris pH 8.0, 600 mM NaCl (elution buffer). The eluted proteins have an apparent high purity in gel (Figure 28B). A new faint 50 kDa band was observed for the BHA_2 sample, presumably a dimer of BHA_2. The BHA_2 dimer may not have bound efficiently to the IE column at pH 6.0 (see *3.2.5* and Figure 26). The composition of the protein samples used in the crystallisation trials was adjusted by dilution with 1% $C_{12}E_{10}$, 25 mM Tris pH 8.0. These samples were: HA_2 10 mg/ml in 1.5% $C_{12}E_{10}$, 25 mM Tris pH 8.0, 300 mM NaCl and BHA_2 9.5 mg/ml in 1.5% $C_{12}E_{10}$, 25 mM Tris pH 8.0, 200 mM NaCl.

The detergent concentration was set to three times of that used for HA, as the molar concentration of HA_2 and BHA_2 samples is about three times higher. The ionic strength in both samples was higher than the aimed 150 mM NaCl, as the obtained protein losses did not allow to dilute the final solutions as much as planned, and the solution concentration could not be adjusted, by dialysis or centrifugation, because $C_{12}E_{10}$ is a non-dialysable detergent.

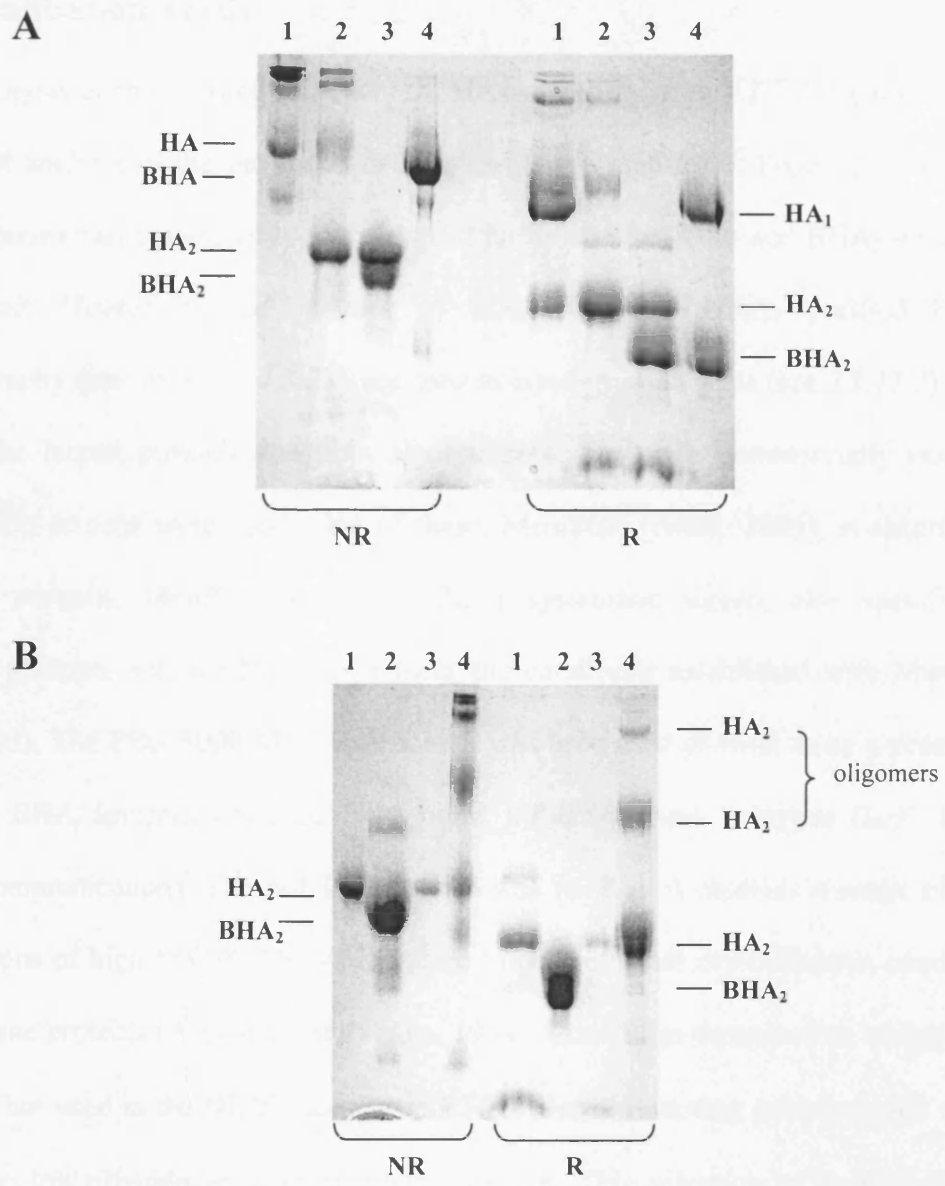


Figure 28. SDS-PAGE of HA and BHA after optimised trypsin digestion, and HA_2 and BHA_2 subsequently purified by IE chromatography

Preparation of HA_2 and BHA_2 for crystallisation trials.

(A) Optimised tryptic digestion of HA and BHA. SDS-PAGE on 12 % gels containing 4 M urea, 0.1 % SDS showing a comparison of the final reaction mixtures for both proteins. Lanes: *1*. Marker of pure HA; *2*. HA trypsin digestion final mixture; *3*. BHA trypsin digestion final mixture; *4*. Marker of pure BHA. R: reduced conditions (0.2% 2-ME). NR: nonreduced conditions.

(B) Purification of BHA_2 (lanes 1 and 2) and HA_2 (lanes 3 and 4) by IE chromatography after optimised tryptic digestion of HA and BHA. SDS-PAGE on 12 % gels containing 4 M urea, 0.1 % SDS showing a comparison of flow-through and elute samples for both proteins. Lanes: *1*. Flow-through; *2*. Elute; *3*. Flow-through; *4*. Elute. R: reduced conditions (0.2% 2-ME). NR: nonreduced conditions.

3.3 Crystallisation Trials

The gel analysis of the purified proteins HA, HA_2 and BHA_2 (see *3.2.7.2*; Figure 28) and the MALDI analysis of the gel bands of HA_2 and BHA_2 (see *3.2.6*; Figure 27) indicated that the proteins had a reasonably high level of purity and that HA_2 and BHA_2 were full-length intact. Therefore, the samples of HA, HA_2 and BHA_2 purified by IE chromatography (see, above, *3.2.7.2*) were used in crystallisation trials (see *2.1.17.2*).

To cover the largest possible spectrum of conditions, the major commercially available crystallisation screens were used. One of these, MemStart (Iwata, 2003), is specific for membrane proteins. MemSys (Iwata, 2003), a systematic screen, also specific for membrane proteins, was used to complement the conditions established with MemStart (Iwata, 2003). The PEG 5000 MME salt screen was used as it showed to be a promising screen for BHA samples obtained from other influenza virus subtypes (L. F. Haire, personal communication). The NR-LBD screen was used as it contains a range of mild concentrations of high MW PEGs, which seem to be preferable crystallisation conditions for membrane proteins (Abramson and Iwata, 1999). Conditions described as having a pH $\leq$ 6.5 were not used in the HDVD assays with HA, to avoid wasting protein which would undergo the low pH-induced conformational change. This selection of conditions was based on final pH values given by the manufacturer (Wizard I and Wizard II screens) or on pH values available in the literature (Wooh et al., 2003), for Crystal Screen (Jancarik and Kim, 1991), Crystal Screen 2 (Jancarik and Kim, 1991; Cudney et al., 1994), Structure Screen 1 (Jancarik and Kim, 1991), and Structure Screen 2 (Jancarik and Kim, 1991; Cudney et al., 1994).

3.3.1 Analysis of crystals

Very thin, 10×300 μm needle-like crystals of HA_2 (data not shown) were grown by MB, in

24%$_{w/v}$ PEG 5000 MME, 0.2 M $MgSO_4$, 0.1 M Tris pH 7.5 and 20%$_{w/v}$ PEG 5000 MME, 0.2 M $MgCl_2$, 0.1 M Tris pH 7.5 (PEG 5000 MME salt screen; see *2.1.17.2*). A few of these crystals were tested using the in-house diffractometer. Very weak Bragg diffraction with closely spaced reflections at low resolution was indicative of protein crystals. However, the solution of the dissolved crystals did not have enough protein to be detected by either EM or Western-blotting. Another very small (icosahedral) crystal was grown by HDVD, but it did not diffract upon testing in the in-house diffractometer. The analysis by EM (carried out by L. J. Calder, at the Division of Virology, MRC-NIMR, London) of the solution (~300 ng μl^{-1} protein concentration) obtained after dissolving the crystal in PBS, showed seemingly trimeric molecules with features of the fusion pH conformation of HA_2, namely globular knobs at the tip of stems (data not shown). The fact that rosettes could not be seen was indicative of the cleavage of the fusion peptide and transmembrane anchor, but it could also indicate that some detergent had remained bound to those regions of HA_2, keeping the molecules soluble. However, the shape of these molecules indicates that HA_2 had been denatured and degraded in significant extent (L. J. Calder, personal communication).

The observed needle crystals had been under crystallisation conditions for ten months, and the icosahedral one for fourteen months. The gel analysis of the BHA_2 sample used in crystallisation trials showed, after twelve months, small MW bands indicative of protein degradation (data not shown). These results suggest that after long periods of time, as the components of the crystallisation drops get highly concentrated, due to water evaporation, very high detergent concentrations may have caused protein denaturation. Also, in those conditions, residual protease activities may become significant. This potential protease activity was not characterised. A residual trypsin activity, used in this study for the preparation of HA_2 and BHA_2, may have contributed for the observed protein degradation.

3.4 Discussion

3.4.1 Isolation and purification in detergent of HA, HA_2 and BHA_2

A protocol for the isolation and purification in detergent of HA, HA_2 and BHA_2 from X-31 influenza virus was prepared in this study. As crystallisation requires a high level of sample homogeneity (Bergfors, 1999), a particular effort was put into having a complete trypsin digestion of low pH-treated HA and BHA, aiming at the preparation of homogeneous solutions of HA_2 and BHA_2. The efficient purification of these two proteins, after trypsin digestion, in a one-single-step purification by IE chromatography ensured not only high yields of protein recovery but also a high sample reproducibility, which is another important factor in protein crystallisation (Bergfors, 1999).

The detergent $C_{12}E_{10}$ (also known as Brij 36T) had proved to be a good dispersing agent for the removal of the full-length HA from the viral envelope (Wharton et al., 1986) and here it also proved to be a suitable detergent to keep HA, BHA, HA_2 and BHA_2 soluble during the low pH incubation required before the trypsin digestion. Also, its mild character indicated that a high detergent concentration necessary to ensure complete protein solubility could be used without risking denaturing the proteins. The concentration of protein samples without concomitant concentration of the nondialysable $C_{12}E_{10}$ was possible by using the IE spin columns, whose large pore (3-5 μm) matrix allowed the free passing of detergent micelles and elution of the matrix-bound protein in the desired volume.

The presence of NA in the protein samples used in the crystallisation assays was not checked. The attempted removal of NA from HA and BHA samples by immunoaffinity chromatography was not completely successful. Some NA was removed by SDG ultracentrifugation and, as it has a pI ≈ 7, at each step of IE. Therefore, the levels of NA in the proteins used in crystallisation trials must be very low. The fact that the crystals which

previously led to the structure determination of viral HA molecules were obtained from samples still containing NA (L. F. Haire, personal communication) suggested that the presence of residual levels of NA in the HA, HA_2 and BHA_2 samples of this study would not be a problem for crystallisation.

3.4.2 Crystallisation trials

The strategy for initial crystallisation trials was to cover a large number of conditions, using different available commercial screens, to increase the probability of getting crystals and identify different crystal forms.

The crystallisation trials were carried out in the detergent $C_{12}E_{10}$, in which the used proteins showed to be stable. The mild character of this detergent suggested that, as the protein solution is driven to high concentrations in the process of crystallisation by HDVD, the risk of protein denaturation might be reduced. However, the EM analysis of HA_2 derived from a crystal (see above) suggests that high $C_{12}E_{10}$ concentrations (among other factors), for long periods of time, may contribute for protein denaturation. The detergents $C_{12}DAO$, also used in this study, and octyl-β-D-glucopyranoside (*β*-OG) are popular detergents in the crystal structures of different membrane proteins (http://www.mpibp-frankfurt.mpg.de/michel/public/memprotstruct.html). However, the proteins under study, after low pH incubation, were unstable in $C_{12}DAO$. Nonetheless, the HA in the neutral pH conformation showed to be stable in this detergent. *β*-OG is an alkyl nonionic detergent, like $C_{12}E_{10}$, having a glycosidic head group instead of the polyoxyethylene of $C_{12}E_{10}$, but a significantly higher concentration of *β*-OG compared to $C_{12}E_{10}$ was required to solubilise full-length HA (data not shown), and it was previously shown that influenza virus envelopes could not be functionally reconstituted after solubilisation with *β*-OG (Stegmann et al., 1987). Also, the majority of the available structures correspond to

membrane integral proteins, which are primarily located in the membrane, and no crystal structure of a full-length membrane-anchored protein similar to the proteins under study has been determined.

3.5 Future Experiments

The crystallisation trials of HA, HA_2 and BHA_2 carried out did not give indications for the improvement of crystallisation conditions. The general aim was to set up an initial crystallisation trial of conditions, using the first protein samples showing reasonably purity and stability, to see if any promising lead was obtained. Further developments to the purification procedures, sample preparation for crystallisation, and crystallisation screens are necessary to create the conditions for the successful crystallisation of these proteins.

3.5.1 Sample preparation for crystallisation

A detailed biochemical characterisation of the protein samples used in crystallisation assays was not done. The ability of the detergent $C_{12}E_{10}$ to completely solubilise the proteins under study, without the formation of non-specific soluble aggregates, and maintain them as trimers, was not tested. Also, the confirmation that HA_2 and BHA_2 were in the low pH-induced conformation was not done. Nonetheless, the EM analysis of the HA_2 dissolved from an obtained crystal in this study (see, above, *3.3.1*) indicates that the molecule was in the trimeric form, even though degraded.

Different other detergents could be tested in their ability to be good solubilisation agents of HA, HA_2 and BHA_2. The detergent screening of the protein samples to be used in crystallisation assays could lead to the selection of other detergents (besides $C_{12}E_{10}$) with different physical properties that could facilitate the crystallisation of the proteins under study. A first approach would be to use two homologues of each detergent class as

previously suggested (Reiss-Husson, 1992). The detergent $C_{12}E_{10}$ could be easily exchanged for other detergents in a single step of IE chromatography.

For the biochemical characterisation of the purified HA, HA_2 and BHA_2 samples, different methods could be implemented. Chemical cross-linking could be used to show that HA_2 and BHA_2 remain trimeric after the purification procedures. The confirmation of the trimeric state of HA_2 and BHA_2 and possible detection of soluble aggregates, checking for size polydispersity, could be done by size-exclusion chromatography and sedimentation equilibrium analytical centrifugation. To ascertain whether HA_2 and BHA_2 were in the fusion pH conformation, mAbs specific for the low pH-induced form of HA_2 could be used, and the formation of HA_2-mAB and BHA_2-mAB complexes checked by EM. The just described methodologies were used in a recent study, for the characterisation of the *E. coli*-expressed full-length HA_2 construct MHA_2, which has the maltose binding protein fused at its N-terminus (Swalley et al., 2004). Dynamic light scattering could also be used to check for size polydispersity of the protein samples included in crystallisation trials (Bergfors, 1999).

The analyses of HA_2 dissolved from a crystal and the BHA_2 solution used in crystallisation assays have shown that after several months the two proteins had undergone degradation (see 3.3.1). The MALDI spectrometry analysis of these two proteins before setting up crystallisation trials indicated that both proteins were full-length intact. This potential protease activity was not characterised. In future experiments, protease inhibitors could be added to the protein samples or used as additives in the crystallisation trials.

3.5.2 Crystallisation trials

Several different crystallisation methods or techniques could be used to screen for new

conditions or improve existent ones. The following are examples of new approaches with more direct application to the crystallisation of HA, HA_2 and BHA_2.

As mentioned above (*3.5.1*), different detergents could be tested in their ability to solubilise HA, HA_2 and BHA_2. In the crystallisation trials, the influence of each detergent on crystallisation could be optimised using small amphiphiles or commercially available detergent screens to create mixed micelles. The characteristics of the detergent micelles have a crucial role on the crystallisation process. Their size or interactions among themselves or with the protein are essential for the structure of the crystal lattice. Small amphiphiles are usually used to reduce the size of the detergent micelle so it can better fit in the protein lattice (Abramson and Iwata, 1999). Preliminary low-resolution crystals of full-length H16 HA in the detergent C_{12}DAO were obtained in the laboratory. The optimisation of the conditions in which these crystals were grown may give valuable indications for the crystallisation of the X-31 HA of this study.

Fab fragments derived from available anti-HA mAbs could be used to mediate the formation of crystals of HA-Fabs complexes, which may be more prone to crystallisation than the mixture of protein and detergent micelles. The detergent molecules, by covering the hydrophobic regions of the protein (the transmembrane anchor and fusion peptide of HA, in this study), may also prevent the formation of essential protein-protein contacts for the assembly of the crystal lattice. Fab antibody fragments can be used to extend the hydrophilic surface of the proteins and create space for detergent micelles, facilitating crystallisation (Hunte and Michel, 2002). For this purpose, Fab fragments of available mAbs able to recognise different regions of the virus native HA, or the HA_2 and BHA_2 in the low pH-induced conformation, could be used to induce the formation of crystals of protein-Fab complexes. Fab fragments directed against the fusion peptide or

transmembrane anchor of HA could also help circumventing the non-specific aggregation problems posed by those regions. Preliminary low-resolution crystals of a NA-Fab complex obtained in the laboratory, promise the applicability of this approach in the crystallisation of HA, HA_2 or BHA_2 in complex with Fab fragments.

Lipidic cubic phases, an alternative crystallisation technique by which the proteins are put under crystallisation conditions after being transferred from the solubilising detergent into a lipid bilayer (Chiu et al., 2000), could also be used. By this method, the proteins to be crystallised would be under the conditions that better mimic those found on the viral envelope.

3.5.3 Other HA proteins

The MAY HA is a non-palmitoylated HA having the three HA_2 C-terminal cysteine residues replaced by methionine, alanine and tyrosine (available in the laboratory). The absence of the palmitoylation modification should decrease the hydrophobicity of both the neutral pH and low pH-induced conformations of X-31 HA, which could make easier the crystallisation of MAY HA and MAY HA_2. These proteins could be prepared using a similar protocol to the one developed for the X-31 HA and X-31 HA_2 of this study. Their structure could also reveal possible implications of the palmitate residues for the structure of HA and HA_2.

4. X-ray Crystal Structure of L226Q HA in Complex with LSTa and LSTc

The structure of L226Q HA in complex with 3'SL [Neu5Acα2-3Galβ1-4Glc], was previously determined (Weis et al., 1988; 1990). Only the Sia moiety of the bound 3'SL could be seen in the structure. Thus, the interactions of the L226Q RBS with the Sia-Gal linkage atoms, a determinant for the understanding of receptor binding specificity, are not known. Here, the receptor analogues LSTa and LSTc (see *Introduction, 1.6.7.1*), whose structures in complex with HAs of different hosts and subtypes have been determined, were used for crystal structure determination in complex with L226Q HA. The L226Q BHA was released from the X-31/HS influenza virus envelope by bromelain digestion, purified and submitted to crystallisation conditions. The crystals obtained were then soaked with the receptor analogues LSTa and LSTc. The BHA form of the different HAs has been used in structural studies of receptor binding (see *1.6.7* and *1.6.2.1.2*) due to the non-specific aggregation problems posed by the full-length molecule (see *1.5.2*, *1.7.5* and *3.*).

4.1 Purification of L226Q BHA

The L226Q BHA was isolated from influenza virus X-31/HS by bromelain cleavage and preliminarily purified by SDG ultracentrifugation, as described for the X-31 BHA in *3.1.2* and *3.1.3*, and then passed through an IE spin column (see *2.1.13*). In this final IE chromatography step of the purification procedure, the spin columns were pre-equilibrated and washed with 25 mM Tris pH 8.0 (loading and washing buffer) and the L226Q BHA was eluted with 25 mM Tris pH 8.0, 300 mM NaCl (elution buffer). The final protein

samples used in crystallisation trials were set to 10 and 38 mg/ml in 10 mM HEPES pH 7.5, 150 mM NaCl, 0.1% azide by centrifugation (see *2.1.10*). In the latter procedure, several steps of sample concentration were carried out, exchanging the IE elution buffer by the crystallisation sample buffer. This buffer exchange was done by diluting the protein with 10 mM HEPES pH 7.5, 150 mM NaCl, 0.1% azide before each centrifugation step. The SDS-PAGE gel analysis of the L226Q BHA sample used in the crystallisation trials indicated that the protein was highly pure (Figure 29).

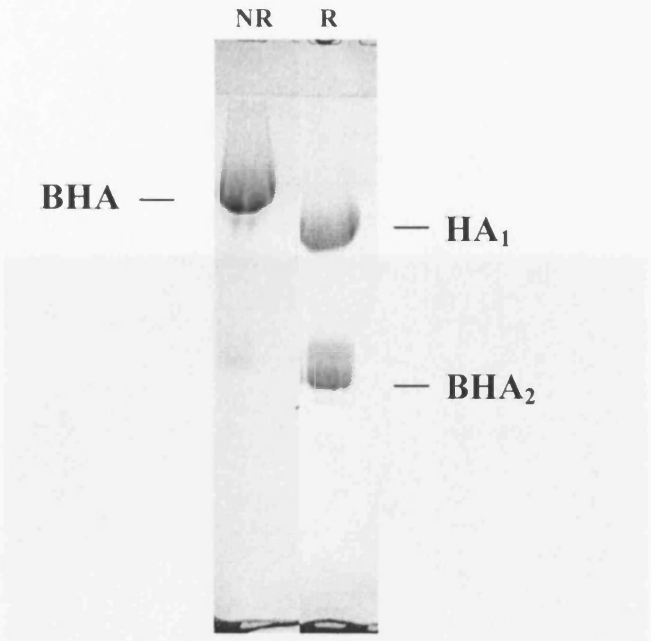
4.2 Crystallisation of L226Q BHA and Soakings with LSTa and LSTc

Screening around the crystallisation conditions previously described for L226Q BHA (Weis et al., 1988) did not produce crystals. Several commercial screens and salt grid screens were then used (see *2.1.17.2*). Diffraction quality crystals (Figure 30) of purified L226Q BHA (see, above, *4.1*) were grown, from a 38 mg/ml protein solution, in 2.1 M tri-ammonium citrate pH 7.5 (see *2.1.17.2*). Crystals mounted from these conditions and soaked with 1.5 mM LSTa or LSTc (see *2.2.1*) for 6 mins were used for data collection. The screening of the conditions for the soaking of the L226Q BHA crystals with LSTa and LSTc revealed that higher concentrations of the ligands or longer incubation times would damage the crystals (an approximate estimate indicated the presence of an average ratio of 750 ligand molecules per L226Q HA RBS during the soaking experiments).

4.3 Data Collection and Processing

The crystals of L226Q BHA soaked with LSTa or LSTc (see, above, *4.2*) diffracted to the resolution of 2.8 Å and 2.9 Å, respectively, at the Daresbury synchrotron.

The data collection (see *2.2.2*) was carried out with as much redundancy as possible, consistent with crystal damage, to improve the data quality. L226Q BHA crystals used for



NR – non reduced conditions
R – reduced conditions (0.2% 2-ME)

Figure 29. SDS-PAGE of L226Q HA purified by IE chromatography

The figure shows the SDS-PAGE analysis on 12 % gels containing 4 M urea, 0.1 % SDS of the L226Q BHA used in the crystallisation trials after purification by IE chromatography. The protein had an apparent high purity.

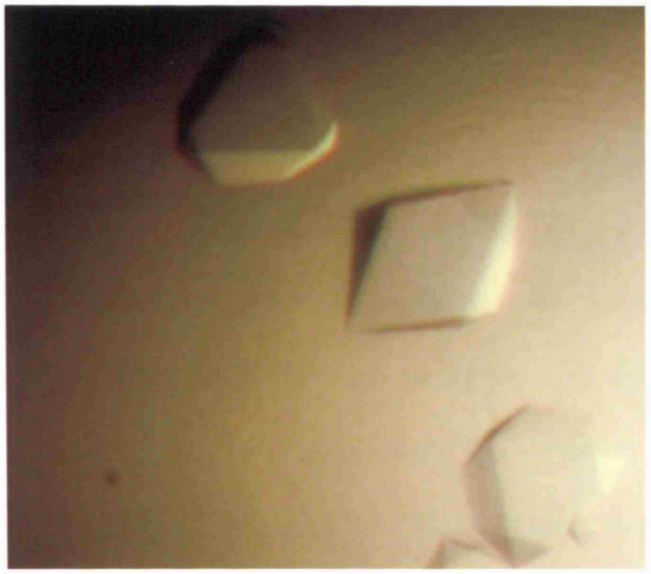


Figure 30. Crystals of L226Q HA used for soaking with the receptor analogues LSTa and LSTc

Crystals of L226Q BHA were grown in 2.1 M triammonium citrate pH 7.5. For soakings, the crystals were harvested in the reservoir solution containing 10% glycerol (cryo buffer) and 1.5 mM LSTa or LSTc. The crystals shown had approximate dimensions of 240x200x40 μm.

soakings with either LSTa or LSTc crystallised near isomorphously in the space group $P4_1$, with one HA trimer in the asymmetric unit and the unusually high solvent content of 78%, typical of $P4_1$ HA crystals (Weis et al., 1990).

Data collection details (see *2.2.2*) and processing statistics (see *2.2.3*) are shown in Tables 2 and 3, for L226Q HA soaked with LSTa and LSTc, respectively.

4.4 Molecular Replacement

The structure of L226Q HA in complex with LSTa and LSTc was solved by molecular replacement (see *2.2.4.2.1*) using the previously determined structure of L226Q HA in complex with 3'SL (Weis et al., 1990) as the search model (see *2.2.4.2.1.1*). The final solutions for molecular replacement, generated by the program AMoRe (Navaza, 2001), are shown in Table 4. The values for the Eulerian angles (α, β, γ), determined by running the cross-Rotation Function, reflect the rotation of the model, which had been rotated in the previous step into the minimal box (see *2.2.4.2.1*), to fit the experimental data. These values are very similar for the complexes L226Q HA-LSTa and L226Q HA-LSTc, since the crystals of both complexes are highly isomorphous. A value of zero was attributed for the fractional translational shift along the *z* axis, since in the space group $P4_1$, in which the L226Q HA has crystallised, there is no fixed origin along *z*. The correct solutions obtained for the complexes L226Q HA-LSTa and L226Q HA-LSTc showed very favourable indicators, with a high correlation value (CC_F) and a relatively low R-factor. The values for both indicators were further improved in the last step of rigid body refinement, in AMoRe (Navaza, 2001).

Table 2 Data collection and processing statistics for L226Q HA in complex with LSTa	
Data collection site	Daresbury SRS
Beamline	Station 14.2
Detector	ADSC Q4 CCD
Detector format	Ccd adsc unsupported-q4
Wavelength (Å)	0.977
Space group	$P4_1$
Unit cell a, b, c (Å) α, β, γ (°)	 160.8 160.8 177.1 90.0 90.0 90.0
Total φ (°), oscillation range (°)	205, 0.50
Mosaicity (°)	0.70
N° total reflections	886060
N° unique reflections	109684 (13608)
Resolution (Å)	2.8 (2.93)
R_{sym} or R_{merge} (%)	11.6 (58.7)
$I/\sigma I$	17.0 (2.9)
Completeness	99.9 (99.3)
Redundancy	8.1 (5.0)

[Values between brackets refer to the highest resolution shell of data processing]

Table 3

Data collection and processing statistics for L226Q HA in complex with LSTc

Data collection site	Daresbury SRS
Beamline	Station 10.1
Detector	MAR 225 CCD
Detector format	Ccd unsupported-m225
Wavelength (Å)	1.283
Space group	$P4_1$
Unit cell a, b, c (Å) α, β, γ (°)	 160.2 160.2 176.0 90.0 90.0 90.0
Total φ (°), oscillation range (°)	180, 0.50
Mosaicity (°)	0.50
N° total reflections	375890
N° unique reflections	95342 (11988)
Resolution (Å)	2.9 (3.03)
R_{sym} or R_{merge} (%)	10.6 (45.5)
$I/\sigma I$	10.1 (2.4)
Completeness	97.4 (98.8)
Redundancy	3.9 (3.9)

[Values between brackets refer to the highest resolution shell of data processing]

	Rotation Function			Translation Function			Unrefined		Refined*	
	α	β	γ	x	y	z	CC_F	R_{factor}	CC_F	R_{factor}
L226Q HA-LSTa	53.10	91.53	279.96	0.2649	0.3058	0.0000	76.2	35.1	76.3	33.9
L226Q HA-LSTc	53.08	91.56	279.93	0.2648	0.3056	0.0000	68.1	38.8	73.2	34.7

Table 4: Molecular replacement solutions for L226Q HA in complex with LSTa and LSTc

The final solutions derived in AMoRe (Navaza, 2001) are shown for the complexes L226Q HA-LSTa and L226Q HA-LSTc. The rotational and translational solution parameters, Eulerian angles (α, β, γ) and fractional translational shifts (x, y, z), respectively, are indicated together with the statistics R_{factor} and CC_F, the correlation coefficient between $|F_{obs}|$ and $|F_{calc}|$. * Rigid body refinement

4.5 Structure Refinement

The refinement statistics for L226Q HA in complex with LSTa and LSTc are shown in Tables 5 and 6, respectively.

Refinement at the resolution of 2.8 or 2.9 Å is very difficult, especially in this particular case of a complex of a protein with highly mobile ligands. The refinement procedures are described below, and a brief discussion of the factors that might have contributed to the poor definition of the electron density of the receptor analogues is presented in the next section (see *4.6.2*).

Refinement was carried out with the program REFMAC (Murshudov et al., 1997) (see *2.2.4.4.1*), which performed restrained refinement and generated the coefficients for the calculation of $2F_o$-F_c density maps using the program FFT (see *2.2.4.5*). Several cycles of refinement were carried out in alternation with model inspection and adjustment using the molecular graphics program O (Jones et al., 1991). The coordinates of LSTa and LSTc were transferred to the model of L226Q HA generated by AMoRe (Navaza, 2001), and several cycles of refinement were run. Dictionary entries for the description of the ligands LSTa and LSTc were generated by REFMAC. Rigid body refinement and manual modelling of the ligands in the electron density was carried out in O (Jones et al., 1991) before each REFMAC refinement cycle. Ligand torsion angle libraries were generated for the manual adjustments (see *2.2.4.4.1*). The saccharide residues of each ligand that could fit into the electron density were selected and the convergence of refinement was monitored by inspection, using the program O (Jones et al., 1991), of the fit of the model, particularly the receptor analogues, in the electron density maps and checking the values for R-factor and R-free factor. No water molecules were added to the model, due to the low resolution of the data (see, below, *4.6.2*).

The non-crystallographic symmetry (NCS) of the L226Q HA trimer, in the refined structures of the complexes L226Q HA-LSTa and L226Q HA-LSTc, can be defined by transformation matrices, which give approximate coordinates for one monomer when applied to each of the other two monomers. These transformation matrices (Table 8, Appendix II) were determined using the program LSQMAN (USF suite) (see *2.2.4.4.1*). No NCS restrains were applied during refinement. The previously determined structure of L226Q HA was used as the search model in molecular replacement to solve the structures of the complexes L226Q HA-LSTA and L226Q HA-LSTc studied here (see, above, *4.4*). The initial electron density maps generated for these complexes showed an overall very good fit to the model, and significant refinement was required only for the regions corresponding to the RBSs and receptor analogues (see, below, *4.5*).

Ramachandran plots of the main chain torsion angles (φ,ψ) for the final L226Q HA-LSTa and L226Q HA-LSTc refined models, generated using the program PROCHECK (Laskowski et al., 1993), are shown in Figure 31. Overall, both structures showed a good agreement of the plot distribution of angles with theoretical energy calculations and distributions observed in structures refined to high-resolution. The L226Q HA in the complexes L226Q HA-LSTa and L226Q HA-LSTc has 86% and 87%, respectively, of the residues in the most probable regions. During refinement, residues outside the most favoured regions were adjusted if consistent with the electron density. The conformations of LSTa and LSTc in complex with L226Q HA are analysed, below, in comparison with their conformations in the available structures of both ligands in complex with other HAs (see *4.7*) and discussed in relation to the conformations of sialosides in solution (see *4.8.1.1* and *4.8.2.1*).

Table 5 Refinement statistics for L226Q HA in complex with LSTa	
Resolution (Å)	2.8
$R_{working}$ (%)	20.6
R_{free} (%) [5% of data]	23.6
Number of atoms	
Total	12220
Protein	12158
Ligand	62
Water	0
Average B_{factor} (Å^2)	
Total	38.3
Protein main chain	37.1
Protein side chain	37.6
LSTa (Sia-Gal) *trans*	111.4
LSTa (Sia-Gal) *cis*	75.7
R.m.s.d.	
Bond lengths (Å)	0.018
Bond angles (°)	1.839

Table 6 Refinement statistics for L226Q HA in complex with LSTc	
Resolution (Å)	2.9
$R_{working}$ (%)	22.1
R_{free} (%) [5% of data]	24.5
Number of atoms	
Total	12198
Protein	12158
Ligand	40
Water	0
Average B_{factor} (Å^2)	
Total	39.4
Protein main chain	38.7
Protein side chain	38.8
Sia	91.4
R.m.s.d.	
Bond lengths (Å)	0.008
Bond angles (°)	1.181

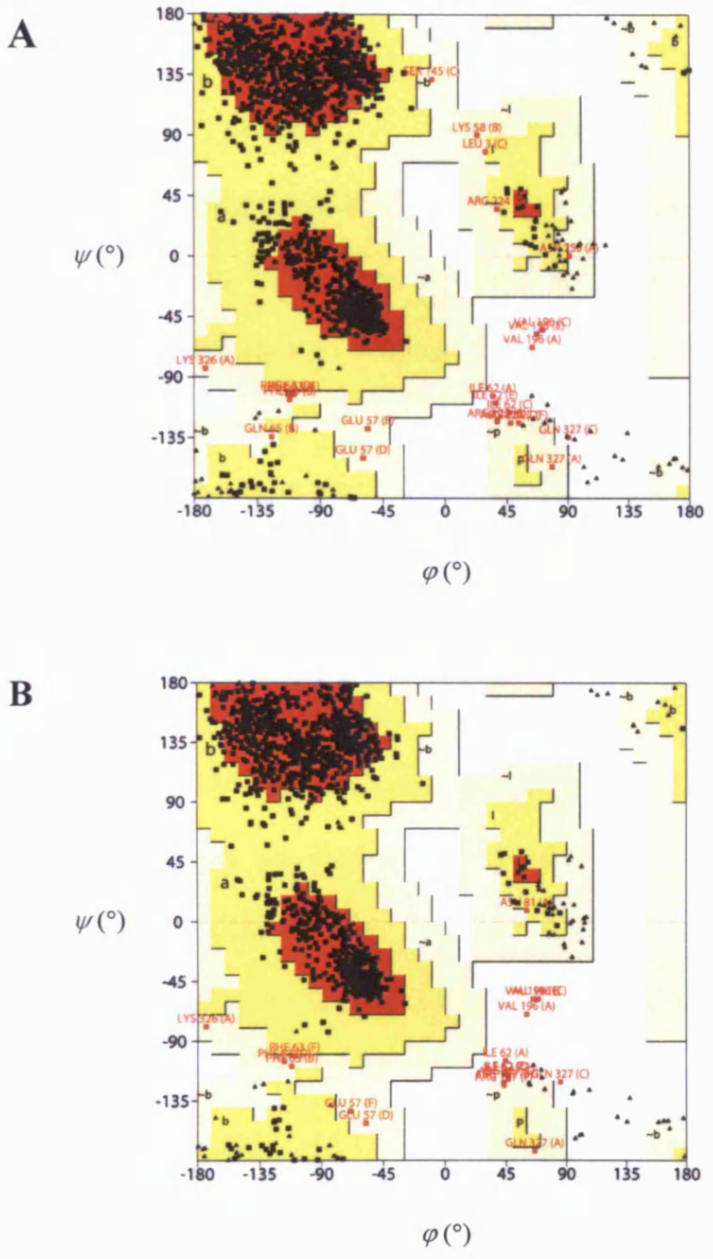


Figure 31. Ramachandran plots for the final refined structures of L226Q HA in complex with LSTa and LSTc

The Ramachandran plots of the main-chain torsion angles (φ, ψ) for L226Q HA in the complexes L226Q HA-LSTa (A) and L226Q HA-LSTc (B) are shown. Triangles represent Gly residues and squares represent all other residues. The most favoured regions (A, B, L) are indicated in red, the additional allowed regions (a, b, l, p) in dark yellow, and the generously allowed regions (~a, ~b, ~l, ~p) in light yellow. Plots were edited from PROCHECK (Laskowski et al., 1993).

4.6 Electron Density Maps for L226Q HA in Complex with LSTa and LSTc

4.6.1 F_o-F_c and $2F_o$-F_c maps

During refinement of the structures of L226Q HA in complex with LSTa or LSTc, F_o-F_c and $2F_o$-F_c electron density maps were generated with the program FFT (see *2.2.4.5*). Poorly defined electron density was obtained for LSTa, up to the second sugar residue, in two monomers of the L226Q BHA trimer (Figure 32). The electron density map at the RBS of L226Q HA soaked with LSTc only allowed the fitting of the Sia residue of the receptor analogue (Figure 33). In both structures, L226Q HA-LSTa and L226Q HA-LSTc, a crystal contact precludes the binding of the receptor analogue to the third HA monomer (see, below, *4.6.3.1*). The calculated omit maps, even though significantly noisy, showed the presence of LSTa and LSTc at the available L226Q HA RBSs (Figures 32A-B, E-F and Figure 33A-B). The refined $2F_o$-F_c maps generated for the complexes L226Q HA-LSTa and L226Q HA-LSTc, including the coordinates of the receptor analogues in the calculation, are shown in Figure 32C-D, G-H and Figure 33C-D. The configuration around the Sia-Gal glycosidic bond of LSTa that could be fitted into the electron density of the two occupied RBSs of a L226Q HA trimer is different between sites. In one site, RBS1, it adopts the *trans* configuration around the Sia-Gal glycosidic bond (Figure 32A-D), and in the other site, RBS2, the *cis* configuration (Figure 32E-H) (see, below, *4.6.3* and *4.7.1.1*). All figures were generated using the program Pymol (DeLano, 2006).

4.6.2 Quality of the electron density

As mentioned above (*4.6.1*), the electron density maps obtained for both LSTa and LSTc in complex with L226Q HA show poor definition, which has made the refinement of the

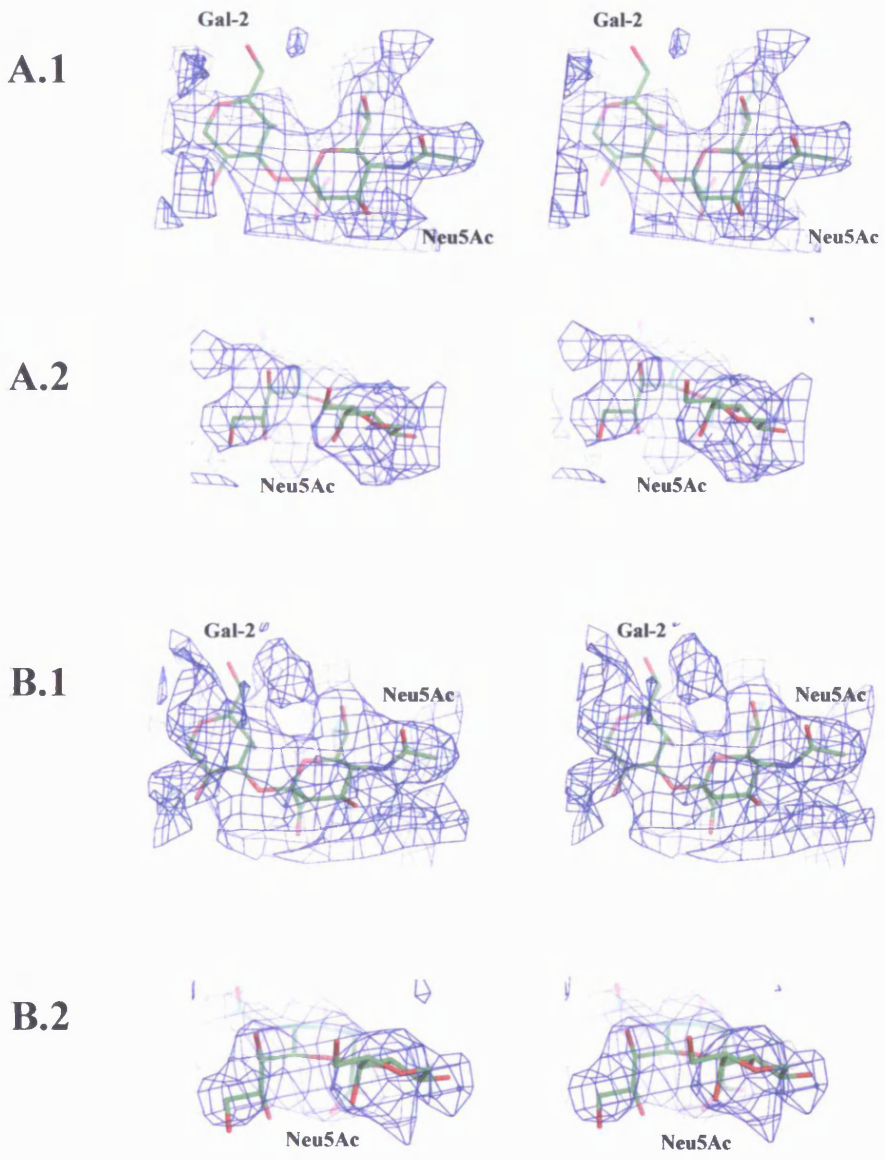
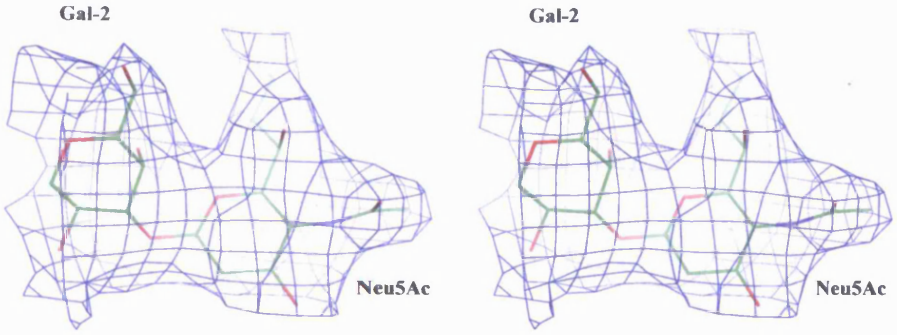


Figure 32. Electron density maps

(A and B). Electron density omit maps of the *trans* conformer of LSTa in complex with L226Q HA

The F_o-F_c and $2F_o$-F_c omit maps shown in (A) and (B), respectively, correspond to the *trans* conformer of LSTa bound to the L226Q HA RBS1 (stereo view). Only Neu5Ac and Gal-2 of LSTa could be fitted into the electron density at the RBS. Maps were selected to cover just the receptor analogue, and in (A.2) and (B.2) the view is focused on the fitting of the 4-OH group of Gal-2 in the electron density. In both maps shown in (A) and (B), the electron density, coloured in blue, is contoured at 0.5 σ level above the mean. Atoms are coloured as: red for oxygen, blue for nitrogen, and green for carbon.

C



D

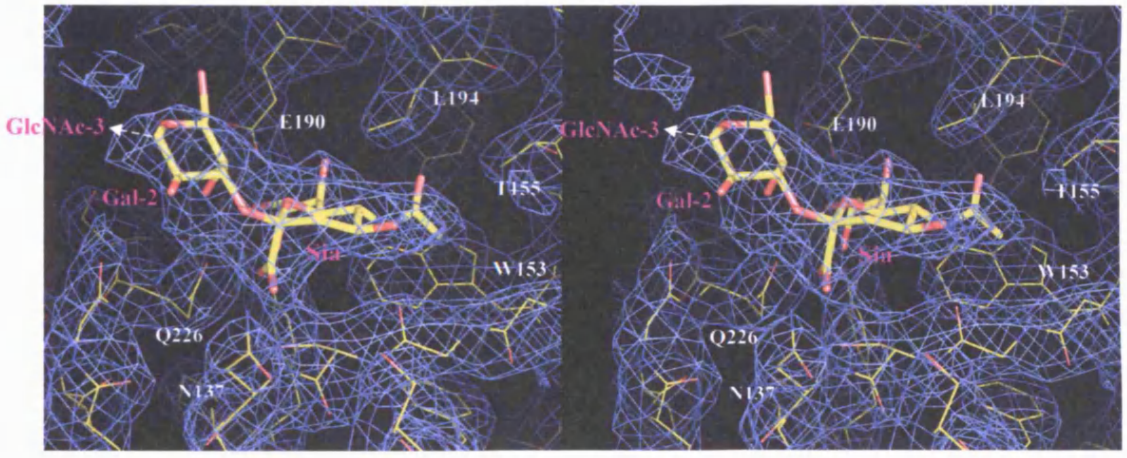


Figure 32 (C and D). Electron density $2F_o$-F_c map of the *trans* conformer of LSTa in complex with L226Q HA

The map shown in (C) and (D) corresponds to the *trans* conformer of LSTa bound to the L226Q HA RBS1 (stereo view). Only Neu5Ac and Gal-2 of LSTa could be fitted into the electron density at the RBS. The map, in (C), was selected to cover just the receptor analogue and, in (D), the electron density corresponding to the residues of the L226Q HA RBS1 is also shown. The electron density, coloured in blue, is contoured at 0.5 σ level above the mean in (C) and 1.0 σ level in (D). Atoms are coloured as: red for oxygen, blue for nitrogen, and green (C) or yellow (D) for carbon.

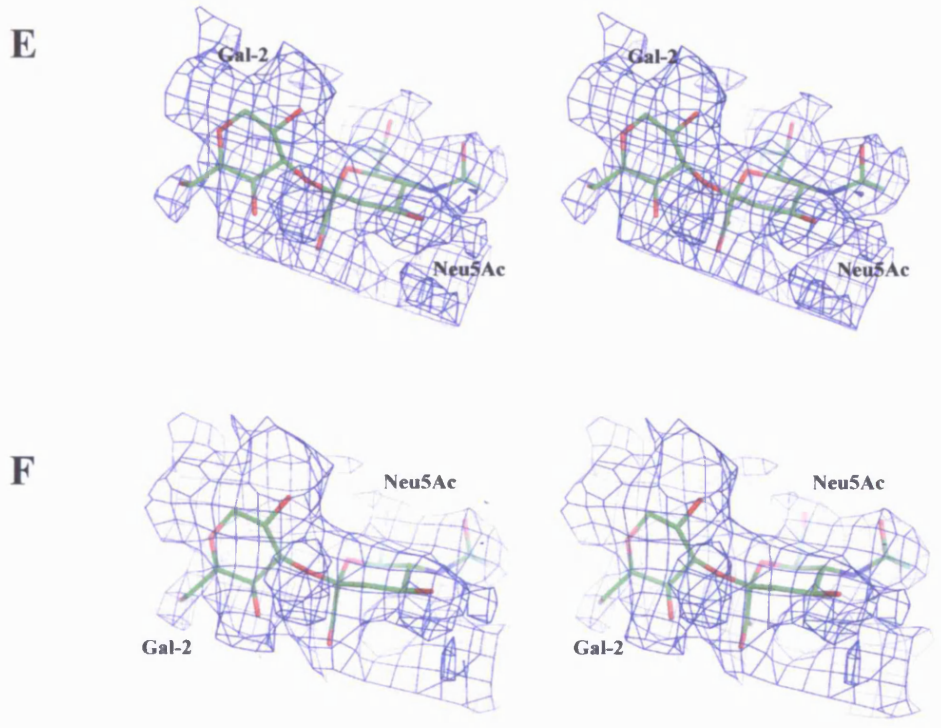
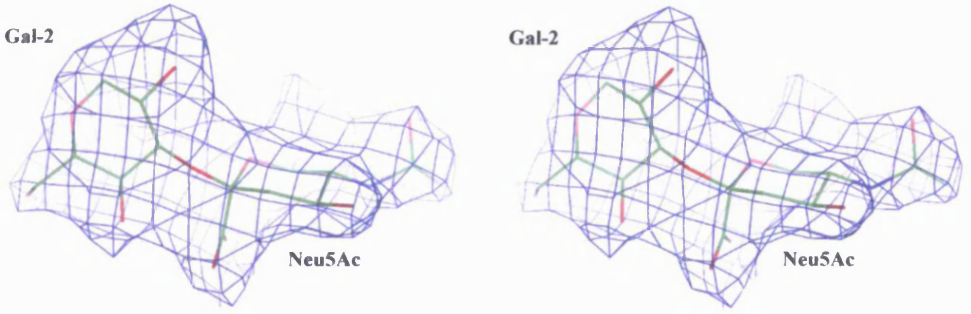


Figure 32 (E and F). Electron density omit maps of the LSTa *cis* conformer in complex with L226Q HA

The F_o-F_c and $2F_o$-F_c omit maps shown in (E) and (F), respectively, correspond to the *cis* conformer of LSTa bound to the L226Q HA RBS2 (stereo view). Only Neu5Ac and Gal-2 of LSTa could be fitted into the electron density at the RBS. Maps were selected to cover just the receptor analogue. The electron density, coloured in blue, is contoured at 0.5 σ level above the mean. Atoms are coloured as: red for oxygen, blue for nitrogen, and green for carbon.

G



H

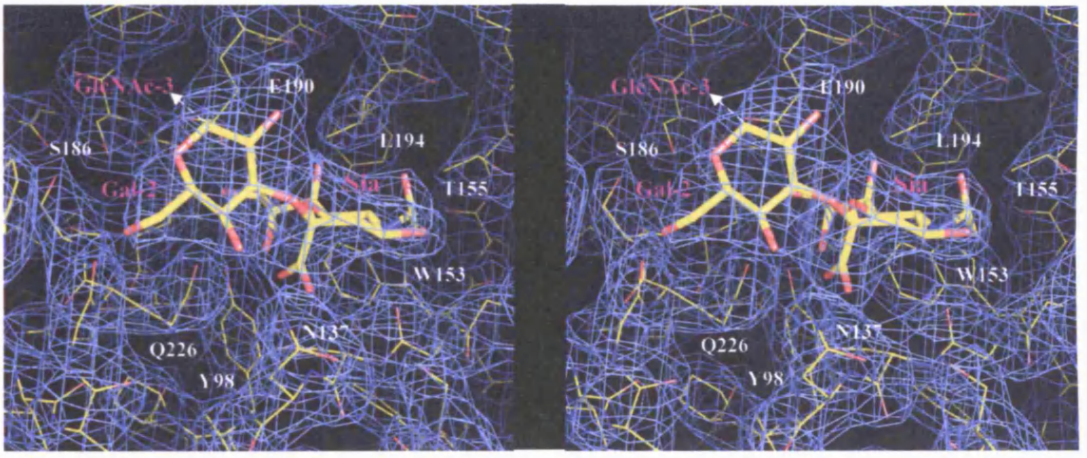


Figure 32 (G and H). Electron density $2F_o$-F_c map of the LSTa *cis* conformer in complex with L226Q HA

The map shown in (G) and (H) corresponds to the *cis* conformer of LSTa bound to the L226Q HA RBS2 (stereo view). Only Neu5Ac and Gal-2 of LSTa could be fitted into the electron density at the RBS. The map, in (G), was selected to cover just the receptor analogue and, in (H), the electron density corresponding to the residues of the L226Q HA RBS2 is also shown. The electron density, coloured in blue, is contoured at 0.5 σ level above the mean. Atoms are coloured as: red for oxygen, blue for nitrogen, and green (G) or yellow (H) for carbon.

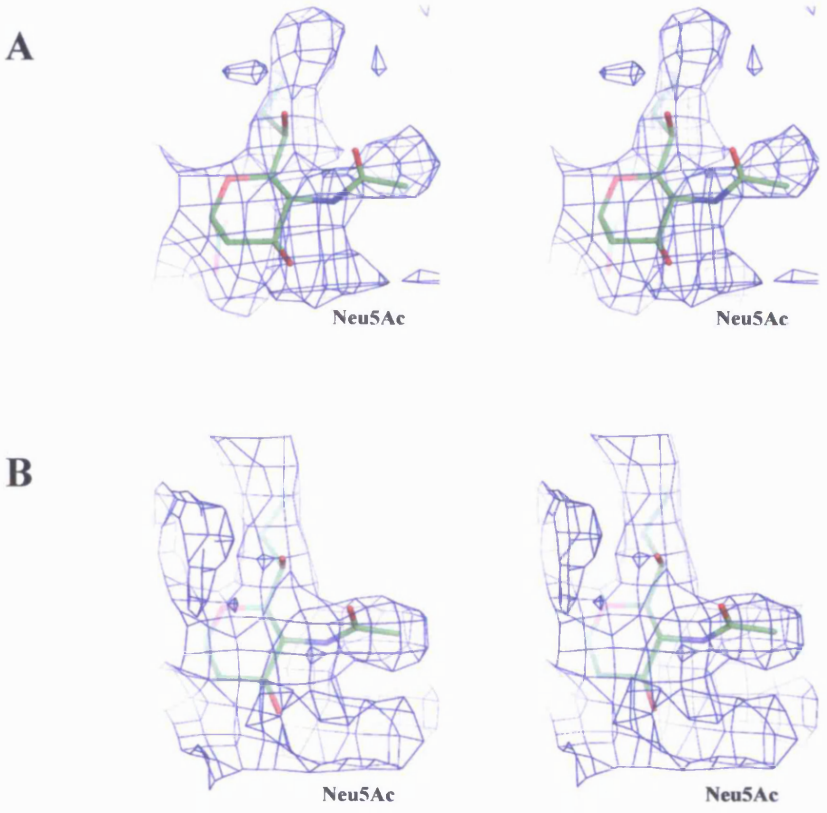
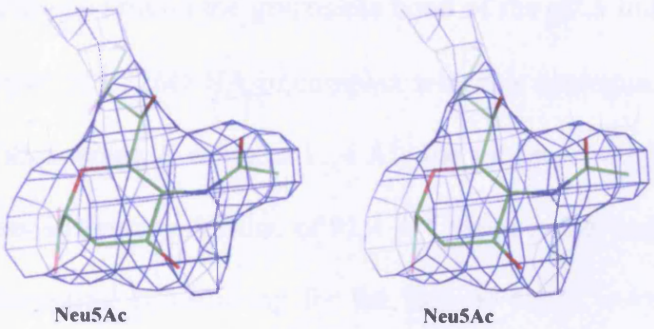


Figure 33. Electron density maps for L226Q HA in complex with LSTc

(A and B). Electron density omit maps of LSTc in complex with the L226Q HA

The F_o-F_c and $2F_o$-F_c omit maps shown in (A) and (B) correspond to LSTc bound to the RBS of L226Q HA (stereo view). Only the Sia of LSTc could be fitted into the electron density. Maps were selected to cover just the receptor analogue. The electron density, coloured in blue, is contoured at 0.7 σ level above the mean. Atoms are coloured as: red for oxygen, blue for nitrogen, and green for carbon.

C



D

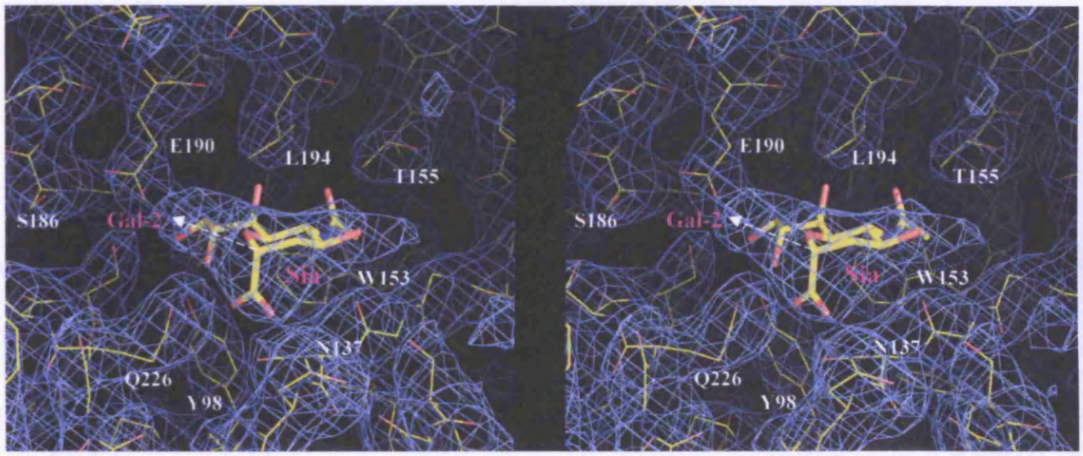


Figure 33 (C and D). Electron density $2F_o$-F_c map of LSTc in complex with the L226Q HA

The map shown in (C) and (D) corresponds LSTc bound to the RBS of L226Q HA (stereo view). Only the Sia of LSTc could be fit in the electron density at the RBS. The map, in (C), was selected to cover just the receptor analogue and, in (D), the electron density corresponding to the residues of the L226Q HA RBS is also shown. The electron density, coloured in blue, is contoured at 0.5 σ level above the mean. Atoms are coloured as: red for oxygen, blue for nitrogen, and green (C) or yellow (D) for carbon.

atomic positional parameters of the receptor analogues very difficult. The overall shape of the saccharide residues is observed in the density, and their general position can be reasonably determined, but the contour for exocyclic groups is poorly defined. Both the *trans* and *cis* configurations around the glycosidic bond of the α2,3 linkage of LSTa were observed in the structure of L226Q HA in complex with this analogue. LSTa in the *trans* configuration shows an average *B* value of 11.4 $Å^2$, and in the *cis* configuration, 75.7 $Å^2$. The Sia of of LSTc has an average *B* value of 91.4 $Å^2$. These values indicate a high degree of disorder of the analogues, contributing for the lack of detail in the electron density. Also, the binding to each RBS of different conformers of the receptor analogues in solution may have contributed for the observed poorly defined density. Another factor influencing the density might be the poor occupancy of the receptor analogues in the crystal.

In contrast with the ligands, the electron density for the protein is of good quality, except for the N-terminus of HA_1 (see below) and a few residues at the surface of the molecule, having high *B* values. The side chains of the RBS residues of L226Q HA in complex with LSTa and LSTc have *B* values ranging from 30 to 60 $Å^2$.

No water molecules were added to the structures of L226Q BHA in complex with LSTa or LSTc, due to low-resolution restrictions that render refinement calculations very uncertain. At contour 0.5 σ there are small peaks in the density however whose position is consistent with water molecules. These are positioned in many cases to make favourable hydrogen bonds. Some of these possible water molecules lie between the receptor analogue and nearby protein, indicating that they could be a factor in the stability of the LSTa-HA or LSTc-HA complexes.

4.6.3 Differences in lattice contacts between non-crystallographically related RBSs

In the isomorphous crystals of X-31 HA and L226Q HA, there is one HA trimer per asymmetric unit and the molecular three-fold axis of the HA trimer is a NCS axis. Therefore, in the asymmetric unit, each RBS of the L226Q HA trimer is in a different environment. This accounts for differences in electron density observed in the three L226Q HA RBSs. Also, crystal contacts between symmetry-related molecules in the crystal may cause slight changes in structure and, in the experiments of soaking with soluble ligands of this study, even block the access of the receptor analogues to a particular RBS. The lattice packing of L226Q HA in the crystals soaked with either LSTa or LSTc was inspected, using the program O (Jones et al., 1991), by generating the symmetry-related molecules.

The atomic interactions between BHA trimers related by the crystallographic $P4_1$ symmetry were previously described (Weis et al, 1990). These interactions are conserved in the $P4_1$ lattices of X-31 HA and L226Q HA. There are two different contact regions, two copies of each per trimer (Figure 34). The most extensive contacts involve two HA_2 strands of the five-stranded antiparallel β-sheet near the base of the molecule. This region, in two monomers of the trimer, is in contact with two different symmetry-related molecules. In each of these two other molecules of the lattice, a different HA_1 region of the top of the trimer is involved in the contacts. One region (contact region B, in Figure 34) comprises residues in surface loops of the HA_1 subunit of a monomer, in the surface parallel to the trimer axis (the L226Q HA RBS1, where the LSTa *trans* conformer binds, is in this monomer); the other (contact region A, in Figure 34), residues of surface loops at the top of the trimer, extending from a small loop (residues 144 and 145) adjacent to the 130 loop of the RBS (L226Q HA RBS2, to which the LSTa *cis* conformer binds) to residues near the left-hand side of the RBS of the neighbouring monomer (L226Q HA

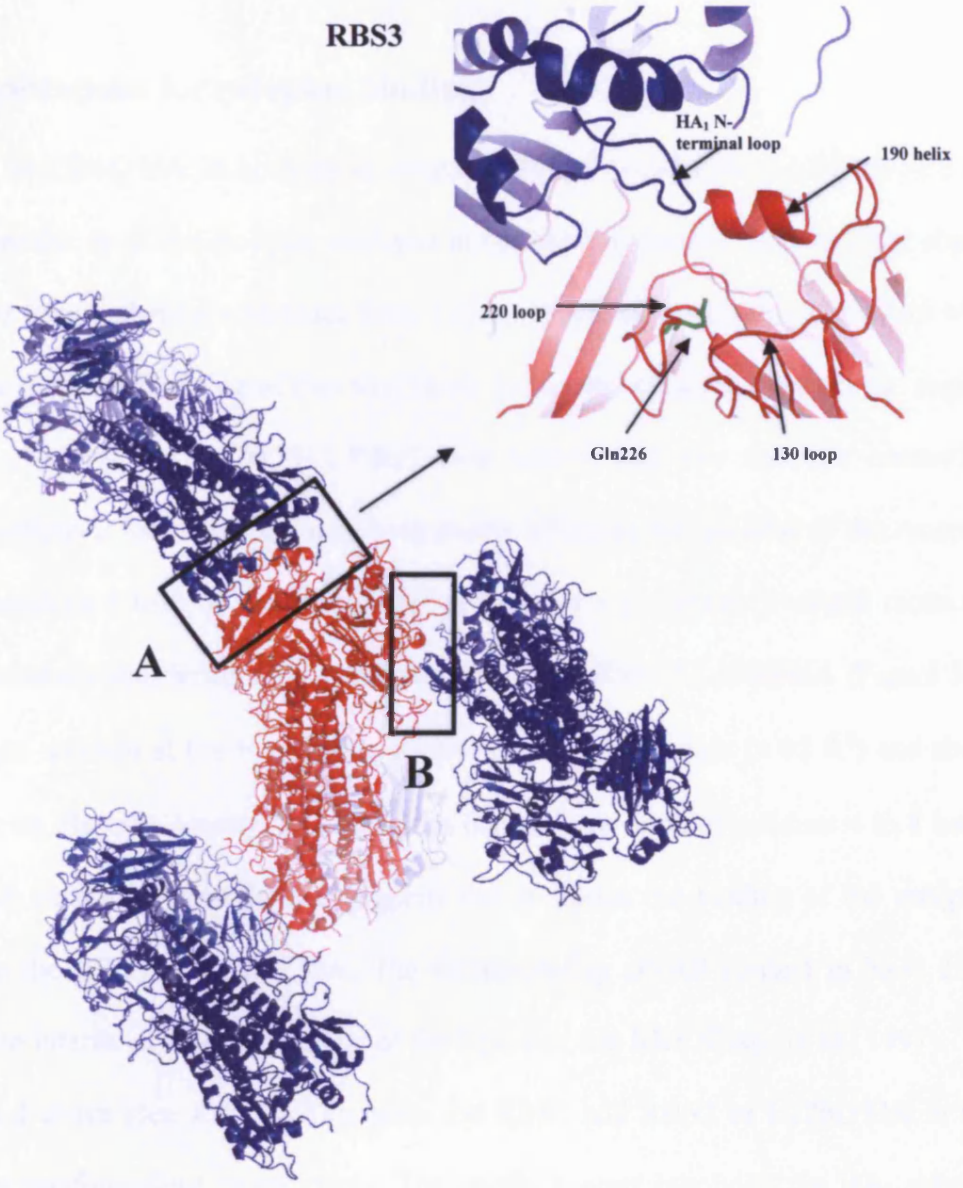


Figure 34. Lattice contacts of the L226Q HA

$P4_1$ lattice contacts of one L226Q trimer (red) with the symmetry-related molecules (blue). The two different regions of the trimeric molecule including residues forming atomic contacts are identified (A and B). Contact region A (magnified at the top of the figure) extends from the right side of RBS2 (to which the LSTa *cis* conformer binds) to the left side of RBS3, which does not bind to either LSTa or LSTc in the crystals of this study. The HA_1 N-terminal loop of the symmetry-related molecule hangs over the site, probably interfering with the binding of the receptor analogue. The side chain of Gln226 in RBS3 is shown (green). The contact region B includes residues of the HA_1 globular head containing the RBS1 to which the *trans* conformer of LSTa binds. The figure was generated using Pymol (DeLano, 2006).

RBS3, showing no bound analogue in this study).

4.6.3.1 Implications for receptor binding

In the RBS3 of L226Q HA, in both the structures L226Q HA-LSTa and L226Q HA-LSTc, it was not possible to fit the receptor analogue in the electron density. As described above (*4.6.3*), in the lattice, there are residues from a symmetry-related molecule in contact with residues near the left-hand side of this site. Here, it was observed that this contact region extends to atoms of the L226Q HA RBS3 (see below) and also that the contacting molecule positions a loop over the site, presumably affecting the binding of the receptor analogue. Residues 4 to 8 of the N-terminus of HA_1 from a symmetry-related molecule form the potentially interfering loop positioned over the RBS3 of L226Q HA (Figure 34). The first eight residues at the N-terminus of HA_1 have high *B* values (> 65 $Å^2$) and show very poor or no electron density. This indicates that the HA_1 loop of residues 4 to 8 has a high level of freedom, which further suggests that it affects the binding of the receptor analogues to the RBS3 of L226Q HA. The corresponding crystal contact in X-31 HA-LSTa has also interfered with the binding of the ligand to that RBS (Eisen et al., 1997).

As mentioned above (see *4.6.1*), LSTa binds the RBS1 and RBS2 of L226Q HA in the *trans* and *cis* configurations, respectively. The crystal contact involving the HA_1 subunit which includes the RBS2 of L226Q HA (contact region A, in Figure 34) might have played a role in the recognition of a different LSTa conformer by that site. The interface areas of the crystal contact regions were determined using the Protein interfaces, surfaces and assemblies (PISA) service at the European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html), authored by E. Krissinel and K. Henrick (Krissinel and Henrick, 2007). The interface area of contact region A is significantly larger than that of contact region B (1006 $Å^2$ and 634 $Å^2$, respectively). In contact region A, 87% (877 $Å^2$)

of the total interface area involves the HA_1 subunit containing the L226Q HA RBS2, which binds the LSTa *cis* conformer. The contact region B only affects residues of the HA_1 subunit harbouring the L226Q HA RBS1, which binds the LSTa *trans* conformer. Therefore, the interface areas of the crystal contacts which might affect the binding preference that the RBS1 and RBS2 of L226Q HA have for the configuration of the Sia-Gal glycosidic bond of LSTa are significantly different, being the contact area for the HA_1 subunit which binds the LSTa *cis* conformer larger than that for the HA_1 subunit binding the LSTa *trans* conformer (877 $Å^2$ and 634 $Å^2$, respectively). The atomic interaction distances in the inter-trimer contact regions in the crystal of the complex L226Q HA-LSTa were determined using the program CONTACT (CCP4 suite) and checked manually using the graphics program O (Jones et al., 1991). The list of these interactions (Table 9, Appendix III) shows that the L226Q HA monomer containing the RBS2, which binds the LSTa *cis* conformer, is involved in more crystal contacts than the monomer harbouring the RBS2, which binds the LSTa *trans* conformer. Furthermore, the contact interface in the RBS2-harbouring HA_1 subunit is contiguous to the RBS2, possibly leading to the formation of a putative small hydrogen bond network: the hydroxylic group of Ser145 from the L226Q HA forms a strong hydrogen bond (3.35 Å) with the carbonyl oxygen of the Asn137 amide, which in turn hydrogen bonds, through its amino group, to the 4-OH of Gal-2 of LSTa (see, below, *4.7.1*). In contact region B, the OH group of Ser145 forms a weak hydrogen bond (3.81 Å) with the side chain of Asn137. Therefore, the interface of crystal contact region A involving the L226Q HA_1 subunit containing the RBS2 might have affected the orientation and contributed to the positional stability of the Asn137 side chain in the RBS2, possibly facilitating the formation of a hydrogen bond between the Asn137 side chain and the Gal residue of LSTa that might have favoured the specific binding of the LSTa *cis* conformer by that site (see *4.7.1*).

4.7 Structure Analysis

The following description and discussion of the structures of L226Q HA in complex with LSTa and LSTc determined in this study are focused on the conformations and atomic interactions that the analogues adopt in the RBSs of L226Q HA. These aspects form a basis for the structural analysis of the receptor binding mechanism of influenza HA.

4.7.1 Structure of L226Q HA in complex with LSTa

4.7.1.1 Configurations of LSTa bound to L226Q HA

The receptor analogue LSTa was fitted in the electron density on two RBS of the L226Q trimer (Figure 32), as described above (*4.6.1*). In the site where the ligand electron density was uninterpretable (RBS3), there was a crystal contact between HA_1 residues adjacent to the RBS and residues of the C-terminus of HA_2 and N-terminus of HA_1 in the membrane-proximal region of a symmetry-related HA molecule (see, above, *4.6.3.1*). A small loop at the N-terminus of the HA_1 subunit of the contacting HA is positioned above the RBS3 of L226Q HA and presumably interferes with the access of LSTa to that particular RBS (Figure 34).

In the two sites where LSTa bound to the L226Q HA, only Sia and Gal could be modelled in the electron density, suggesting that the receptor analogue is disordered beyond Gal-2 (Figure 32A and B). As mentioned in the previous section, the Sia-Gal portion of LSTa adopted two different configurations in two different RBSs of L226Q HA. In each site, a differently located protrusion in the density globule covering the Gal ring was interpreted as corresponding to the C6'H_2OH exocyclic group, with the consequence that in one of the sites, RBS1, LSTa was assigned the *trans* configuration around the glycosidic bond, and in the other, RBS2, the *cis* configuration. Figures 35 and 36 show the *trans* and *cis* configurations, respectively, adopted by LSTa on two different monomers of each L226Q

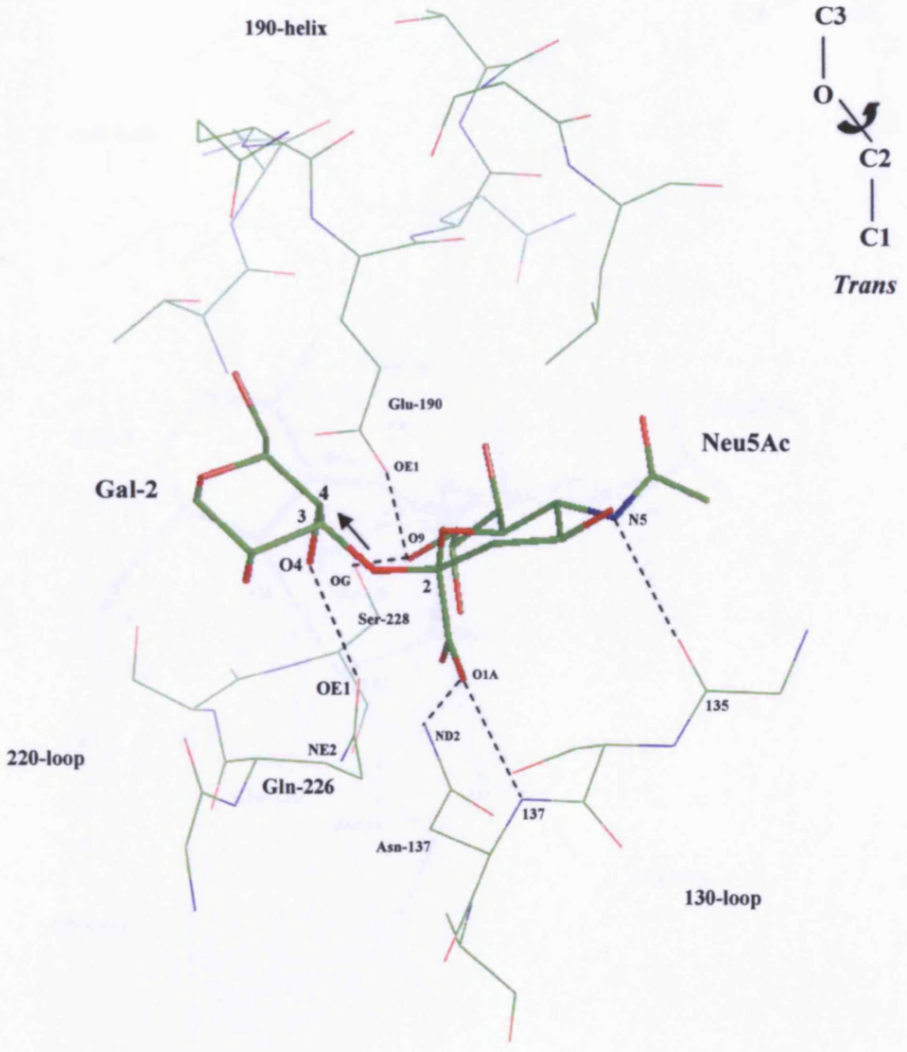


Figure 35. Binding of LSTa in the *trans* configuration to L226Q HA

The *trans* conformer of LSTa binds to the RBS1 of L226Q HA. Selected potential hydrogen bonds between analogue and L226Q HA are shown as dashed lines. Atoms are coloured as: green for carbon, blue for nitrogen, and red for oxygen. Relevant residues of the HA RBS and atoms of LSTa and the HA RBS, namely those involved in the selected interatomic contacts, are specifically identified (see: Figure 40A; *4.7.1.2* and *4.8*; Table 10, Appendix IV). The three main secondary structural elements of the HA RBS - 130 loop, 190 helix and 220 loop - are also indicated. The arrow indicates the configuration around the glycosidic bond, as shown schematically at the top right of the figure.

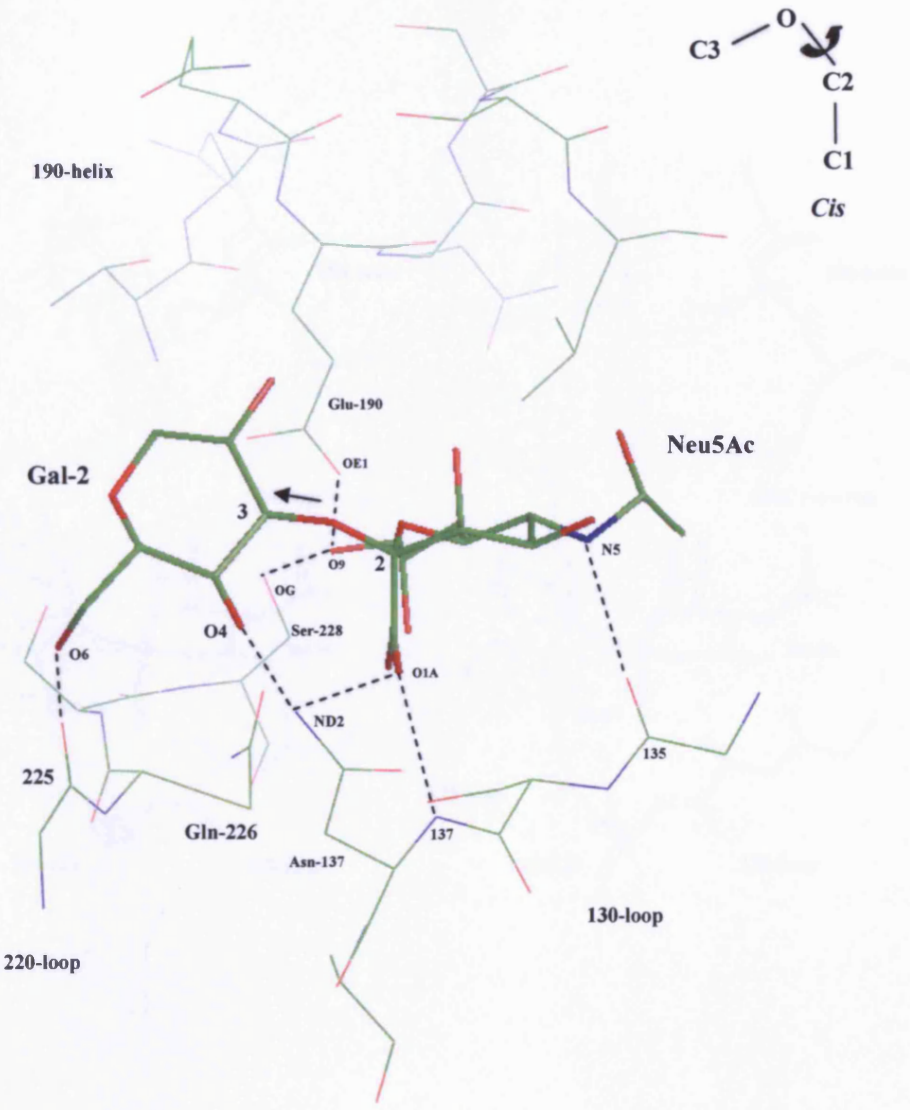


Figure 36. Binding of LSTa in the *cis* configuration to L226Q HA

The *cis* conformer of LSTa binds to the RBS2 of L226Q HA. Selected potential hydrogen bonds between analogue and L226Q HA are shown as dashed lines. Atoms are coloured as: green for carbon, blue for nitrogen, and red for oxygen. Relevant residues of the HA RBS and atoms of LSTa and the HA RBS, namely those involved in the selected interatomic contacts, are specifically identified (see: Figure 40B; *4.7.1.2* and *4.8*; Table 10, Appendix IV). The three main secondary structural elements of the HA RBS - 130 loop, 190 helix and 220 loop - are also indicated. The arrow indicates the configuration around the glycosidic bond, as shown schematically at the top right of the figure.

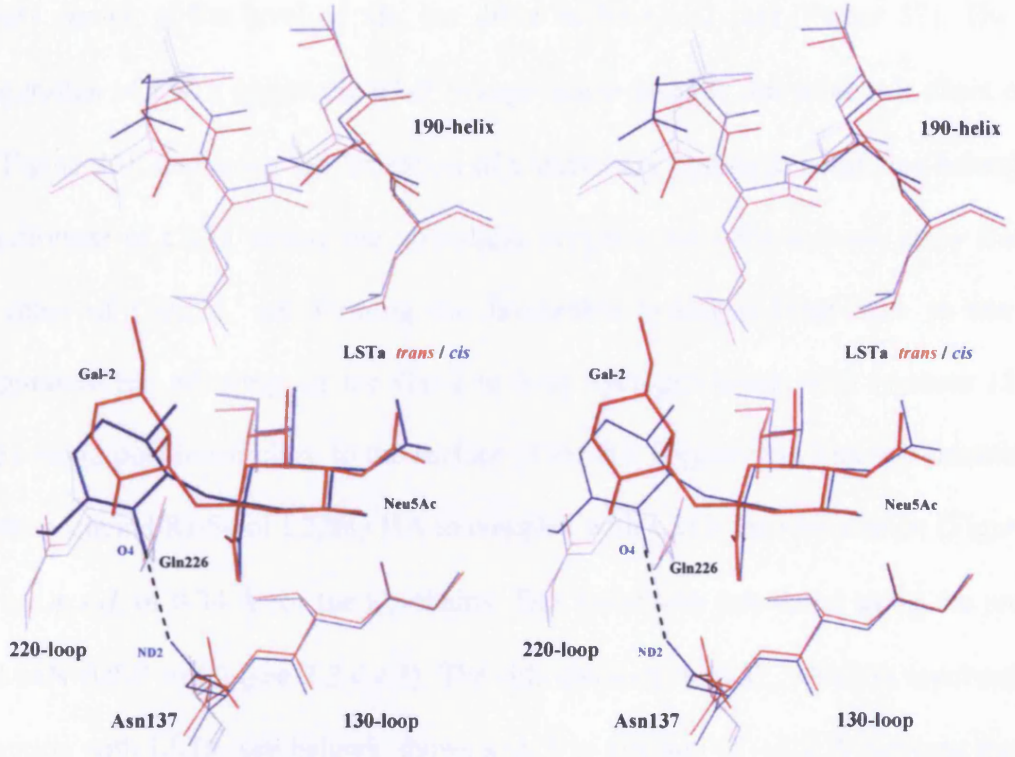


Figure 37. Comparison of the structure of LSTa in the *trans* and *cis* configurations in complex with L226Q HA

The complex of the LSTa *trans* conformer with the L226Q HA RBS1, shown in red, is overlaid on that of the LSTa *cis* conformer with the L226Q HA RBS2, shown in blue (stereo view). The new hydrogen bond formed by the side chain of Asn137 with the 4-OH of Gal-2 of the LSTa *cis* conformer is shown as a dashed line. The three main secondary structural elements of the HA RBS - 130 loop, 190 helix and 220 loop - are also indicated.

trimer, and some of the potential hydrogen bonds between the receptor analogue and the protein (analysed below). In Figure 37, the superposition of both *trans* and *cis* conformers of LSTa bound to the L226Q HA is shown. The structures of the two bound conformers are very similar at the level of Sia, but differ in the Gal-2 part (Figure 37). The *trans* configuration of LSTa points the α2-3 linkage atoms towards the polar side chain of Gln 226 (Figure 35), favouring the formation of a stabilising hydrogen bond (see below). The *cis* conformer of LSTa directs the glycosidic oxygen toward the solvent, away from the side chain of Gln226, not forming the favourable hydrogen bond seen in the *trans* configuration but allowing for the Gal-2 to form hydrogen bonds with residues 137 and 225 by being positioned close to the surface of the HA (Figure 36). The conformations of the two occupied RBSs of L226Q HA in complex with LSTa are very similar (Figure 37), with a *r.m.s.d.* of 0.34 Å for the C_α-chains. This value was calculated using the program LSQMAN (USF suite) (see *2.2.4.4.1*). The side chain of Asn137, which is involved in an interaction with LSTa (see below), shows a shift in position of ~ 1.2 Å between the sites. This distance is significant since the coordinate error in the residue is about 0.4 Å (*r.m.s.d.* of 0.8 Å, derived from the *B* value of the amide nitrogen). Thus the side chain of Asn137 may specifically adopt an optimal positioning in the RBS2 to hydrogen bond to the 4-hydroxyl group of Gal-2 of the *cis* configuration of LSTa (Figure 36).

In the crystal structures of HA-LSTa complexes determined so far (Ha et al., 2001, 2003; Gamblin et al., 2004; Russell et al., 2006), LSTa binds all the avian HAs and the H1 human HA in an extended conformation, and in the *trans* configuration around the Sia-Gal glycosidic bond (Figures 8A, 9A, 10, and 12A). The superposition of the structure of LSTa in the *trans* configuration when bound to the L226Q HA with those when in complex with H3 avian HA, H5 avian HA, H7 avian HA and H1 human HA is shown in Figure 38. The overlaid structures show that LSTa binds the RBS1 of L226Q HA, the

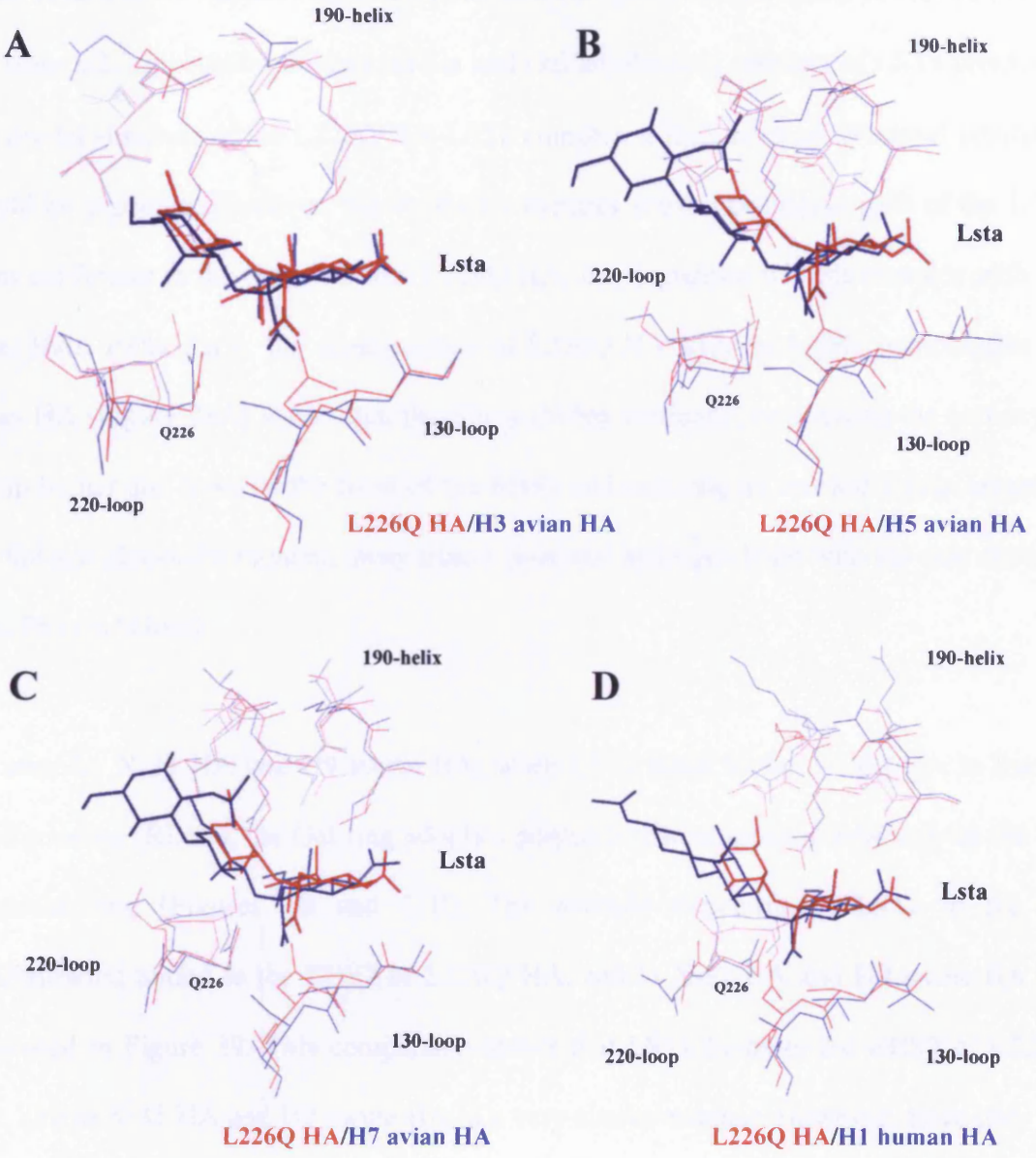


Figure 38. Comparison of the binding of the LSTa *trans* conformer to L226Q HA with that to H3 avian HA, H5 avian HA, H7 avian HA, and H1 human HA

Superposition of the L226Q HA RBS1 and those of the H3 avian HA (A), H5 avian HA (B), H7 avian HA (C), and H1 human HA (D) in complex with LSTa in the *trans* configuration. The HAs are identified by colour. For the H3 avian HA, the RBS where LSTa was ordered only for Sia and Gal was chosen. The other two monomers are bound to the first three saccharide residues of Lsta, as shown in Figure 8A (see *Introduction*). The three main secondary structural elements of the HA RBS - 130 loop, 190 helix and 220 loop - are also indicated.

avian HAs and H1 human HA in a similar manner. Since this comparison was limited to the *trans* α2-3 glycosidic linkage, as Sia and Gal are the only residues of LSTa present in the crystal structure of the L226Q HA-LSTa complex, a high level of structural similarity would be expected. However, the overlaid structures show a positional shift of the LSTa *trans* conformer in the complex with L226Q HA, in comparison with its complex with the other HAs. Particularly, the superposition of L226Q HA with the highly homologous H3 avian HA (Figure 38A) shows that the Sia is shifted vertically, positioning the carboxylic group higher and towards the front of the RBS1 and inducing an upward displacement of the linkage glycosidic oxygen, away from a potential hydrogen bond with the side chain of Gln226 (see below).

As seen for X-31 HA and H9 swine HA, when LSTa binds to the L226Q HA in the *cis* configuration (RBS2), the Gal ring adopts a perpendicular orientation relatively to the Sia pyranose ring (Figures 7B and 11B). The overlaid structures of LSTa in the *cis* configuration bound to the RBS2 of L226Q HA, and to X-31 HA and H9 swine HA are presented in Figure 39. This comparison shows that LSTa binds to the RBS2 of L226Q HA, and to X-31 HA and H9 swine HA in a very similar manner. However, since only the Sia and Gal residues are seen in the structures of the L226Q HA-LSTa and H9 swine HA-LSTa complexes, the comparison was limited to the *cis* α2-3 linkage, and hence a high structural similarity was expected. [Figures 35-39, 41 were generated using Pymol (DeLano, 2006)]

4.7.1.2 Atomic interactions between LSTa and L226Q HA

A schematic representation of the atomic interactions of LSTa in the *trans* configuration with RBS1 of L226Q HA, and LSTa in the *cis* configuration with RBS2 of L226Q HA is shown in Figure 40. For comparison, a list of atomic interaction distances of LSTa with

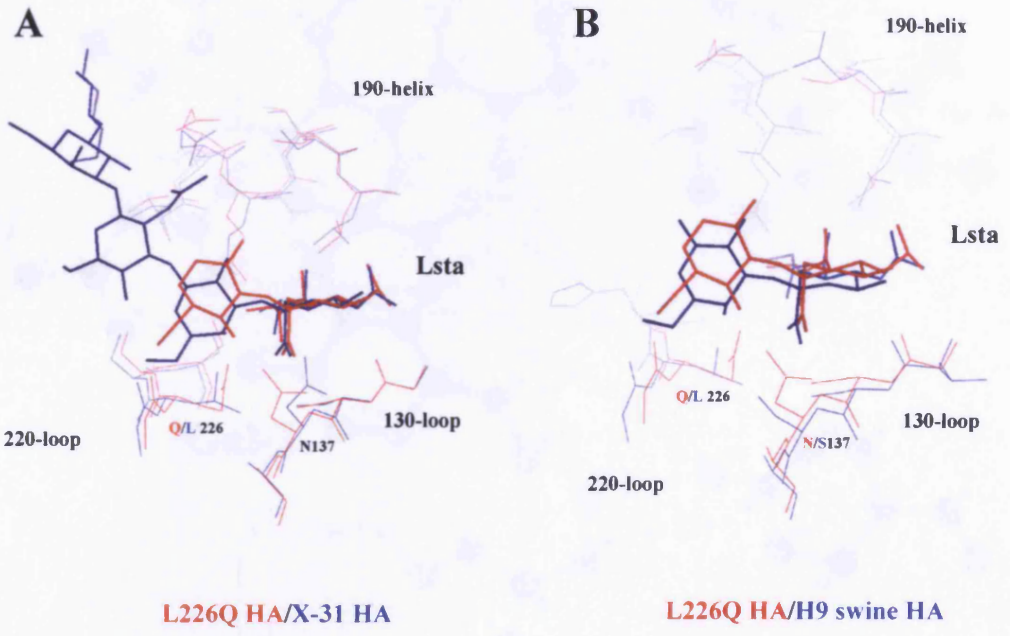


Figure 39. Comparison of the binding of the LSTa *cis* conformer to L226Q HA with that to X-31 HA and H9 swine HA

Superposition of the RBS2 of L226Q HA and the RBSs of X-31 HA (A) and H9 swine HA (B) in complex with LSTa in the *cis* configuration. The HAs and corresponding receptor analogue are identified by colour.

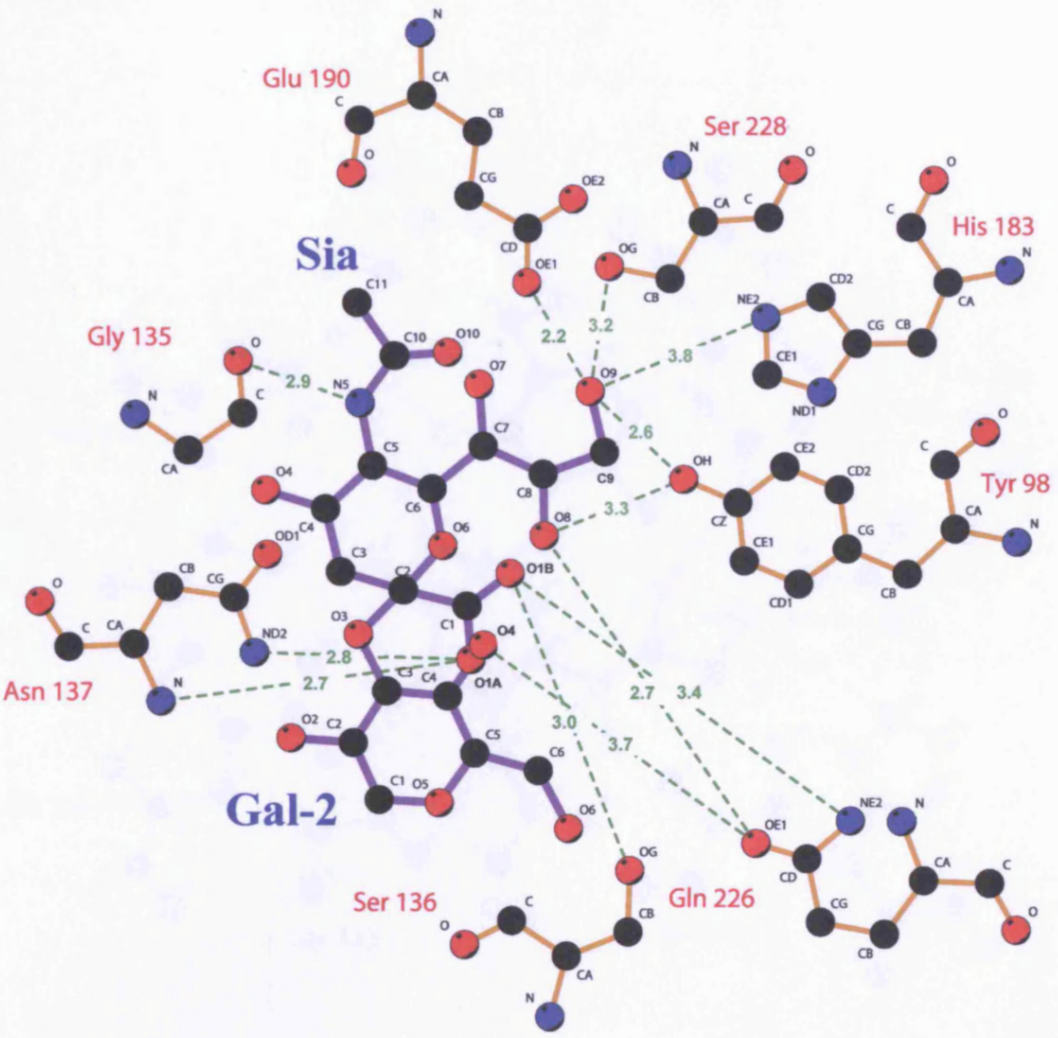


Figure 40. Interactions of LSTa with the L226Q HA

A. Schematic diagram of the interactions of LSTa in the *trans* configuration with the residues of the RBS1 of L226Q HA

Potential hydrogen bonds (shown as dashed lines) between the LSTa *trans* conformer and the residues of the RBS1 of L226Q HA are shown. Only Sia and Gal-2 of LSTa are shown, since these were the only residues of the ligand that could be fitted into the electron density (see *4.6*). The interatomic interaction distances corresponding to the represented hydrogen bonds are indicated (see, also, Table 10, Appendix IV). Atoms (shown as balls) are coloured as: black for carbon, blue for nitrogen, and red for oxygen. Covalent bonds (shown as sticks) in LSTa and the L226Q HA residues are coloured in purple and brown, respectively. The identity of main-chain and side chain atoms is indicated.

The figure was generated using the program LIGPLOT (Wallace et al., 1995).

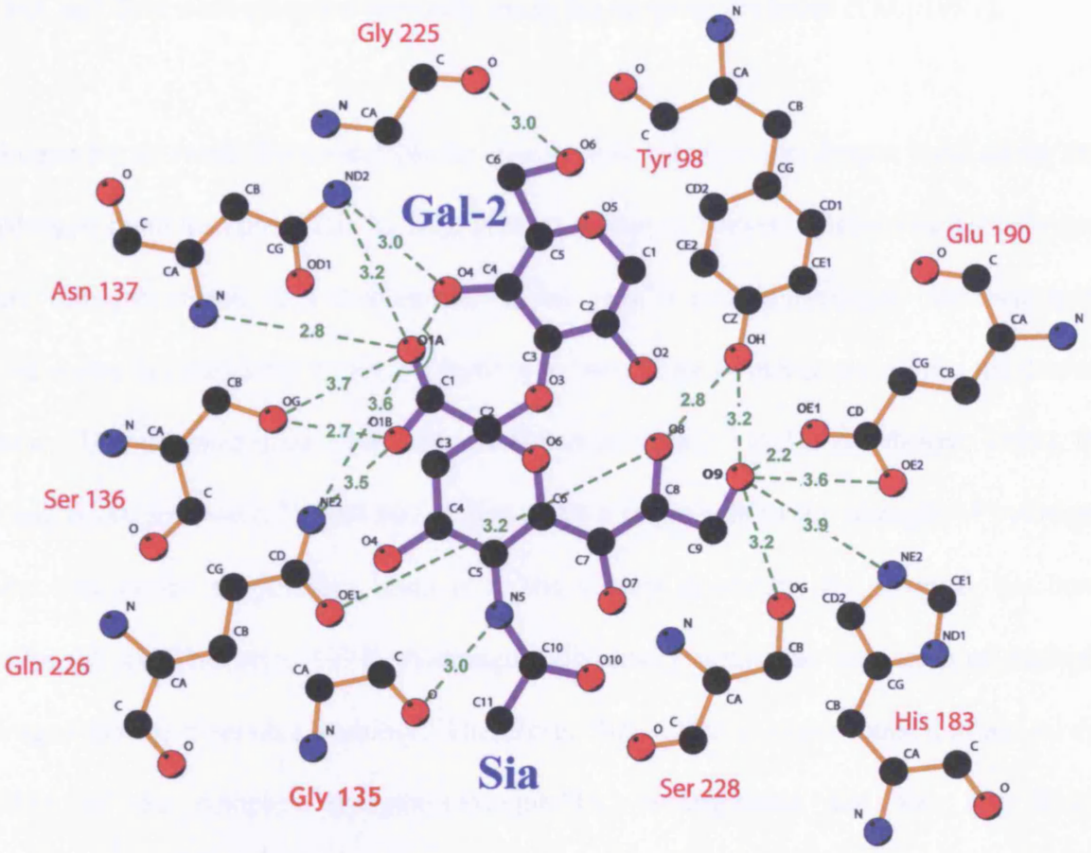


Figure 40B. Schematic diagram of the interactions of LSTa in the *cis* configuration with the residues of the RBS1 of L226Q HA

Potential hydrogen bonds (shown as dashed lines) between the LSTa *cis* conformer and the residues of the RBS2 of L226Q HA are shown. Only Sia and Gal-2 of LSTa are shown, since these were the only residues of the ligand that could be fitted into the electron density (see *4.6*). The interatomic interaction distances corresponding to the represented hydrogen bonds are also indicated (see, also, Table 10, Appendix IV). Atoms (shown as balls) are coloured as: black for carbon, blue for nitrogen, and red for oxygen. Covalent bonds (shown as sticks) in LSTa and the L226Q HA residues are coloured in purple and brown, respectively. The identity of main-chain and side chain atoms is indicated.

The figure was generated using the program LIGPLOT (Wallace et al., 1995).

L226Q HA, X-31 HA (H3 human), H3 avian HA, H5 avian HA, H7 avian HA, H1 swine HA, H9 swine HA and H1 human HA are shown in Table 10, Appendix IV. The interaction distances shown are for potential hydrogen bonds between receptor analogue and HA and they were assigned manually using the program O (Jones et al., 1991).

Hydrogen bonds result from electrostatic interactions between a hydrogen bond donor and a hydrogen bond acceptor. The hydrogen atom is shared between electronegative oxygen atoms, nitrogen atoms, or a combination of the two. A proper hydrogen donor-acceptor pair of atoms is considered to form a hydrogen bond if those atoms are within the correct distance. This distance is generally considered to be from 2.7 to 3.3 Å (McRee, 1999), for a strong hydrogen bond. Directionality also plays a major role in the strength of hydrogen bonds. The closer a hydrogen bond is to the correct geometry, the stronger the bond (McDonald and Thornton, 1994). Although individually weak, the formation of multiple hydrogen bonds promotes stability. Therefore, they offer a major contribution to the stability of the complex receptor-analogue/HA. Hydrophobic and van der Waals interactions are weaker than hydrogen bonds and non-directional. A van der Waals contact is attributed to two atoms if the distance between them is less than 4.0 Å (Weis et al., 1990).

The interaction pattern between the Sia residue of LSTa and the RBS1 and RBS2 of L226Q HA (Figure 40) is very similar to that seen in the structures of HAs of different hosts and subtypes in complex with either LSTa or LSTc (Tables 10 and 11, Appendix IV). This is consistent with the general conservation of the HA – Sia binding pattern, irrespective of the analogue, even in the structures of HA in complex with receptor analogues other than LSTa and LSTc (reviewed in Skehel and Wiley, 2000; see

Introduction, 1.6.7). The structure of the HA_1 globular domain of different HAs is very similar and subsequently the ligand-free RBS region is conserved in large extent between different HAs.

In both configurations of LSTa bound to L226Q HA, the conserved interaction of the 9-hydroxyl group of the glycerol moiety of Sia with the imidazole group of H183 is very weak (close to 4 Å, see Figure 40). The L226Q HA shares with the X-31 HA the ability to form a direct hydrogen bond of the side chain of Ser228 with the 9-hydroxyl group of the glycerol side chain of Sia, and this interaction was observed for both *trans* and *cis* conformers of LSTa in complex with L226Q HA (Figures 7A, 7B, 35, 36; Table 10, Appendix IV). Avian HAs, H9 swine HA and H1 HAs whose structure in complex with LSTa was determined (Ha et al., 2001; Ha et al., 2003; Russell et al., 2006; Gamblin et al., 2004) have Gly at position 228. Thus, these HAs do not directly hydrogen bond to Sia at that position. Nonetheless, in the structures of LSTa bound to H5 avian HA and H9 swine HA, a water molecule mediates the hydrogen bonding between HA residue 228 and Sia (Ha et al., 2001) (Table 10, Appendix IV). The presence of either Ser or Gly at position 228 affects the affinity of receptor binding of H2 and H3 HAs, influencing their specificity for either α2-3 or α2-6 linkages (Naeve et al., 1984; Vines et al., 1998; Connor et al., 1994, Matrosovich et al., 2000). This might be related, in part, with the ability of the HA to form or not a direct hydrogen bond with the Sia of the receptor through the side chain of residue 228.

In the structures of H3 avian HA, H5 avian HA, and H7 avian HA in complex with LSTa (Ha et al., 2001; Ha et al., 2003; Russell et al., 2006) the amino (nitrogen NE2) and carbonyl (oxygen OE1) groups of the amide of the Gln226 side chain form potential hydrogen bonds with the Sia-Gal glycosidic oxygen and the 4-hydroxyl group of Gal-2 of LSTa, respectively (Figures 8A, 9A, and 10; Table 10, Appendix IV). These interactions

are made possible by the *trans* configuration of the α2,3 linkage of LSTa, which positions the glycosidic oxygen directly over the amide group of the Gln226 side chain and projects the Gal-2 ring up and out of the RBS, in such a way that the axial 4-OH group can hydrogen bond to the carbonyl group of the Gln226 side chain (Figures 8A, 9A, and 10). The just described structural complementarity of both the glycosidic oxygen of the Sia-Gal α2,3 linkage and the 4-OH group of Gal-2 of LSTa in the *trans* configuration to the hydrogen bond potential of the side chain of Gln226 defines a recognition motif specific for the *trans* configuration of α2,3 linkages, as fist identified in the crystal structure of the H5 avian HA in complex with LSTa (Ha et al., 2001).

In the complex of the RBS1 of L226Q HA with LSTa in the *trans* configuration (Figure 35), the side chain of Gln226 forms a potential hydrogen bond, via the amide carbonyl group, with the 4-hydroxyl group of Gal-2 (Figure 40A), as seen in other HAs that specifically recognise only the *trans* configuration of the α2,3 linkage (see just above). However, the amino group (NE2) of the side chain of Gln226 is out of hydrogen bond range with the Sia-Gal glycosidic oxygen (5.5 Å). This shows that the L226Q HA, in contrast with the other HAs which form a recognition motif specific for the *trans* configuration of the α2-3 linkage (see just above), is not able to form a hydrogen bond with the glycosidic oxygen of the linkage. As described in *4.6.1.1* a displacement of the LSTa *trans* conformer (Figure 38) has caused a shift in the position of the glycosidic oxygen, which is positioned too far away from the side chain of Gln226 to form a hydrogen bond.

The main-chain carbonyl of residue 225 of L226Q HA forms a potential hydrogen bond with the exocyclic 6-hydroxyl group of Gal-2 of the *cis* conformer of LSTa (Figures 36 and 40B). The same interaction is observed in the complexes of X-31 HA and H9 swine

HA with LSTa (Figures 7A and 11A; Table 11, Appendix IV), which binds these HAs in the *cis* configuration, and it is the only interaction that those two HAs form with the asialo portion of LSTa. Therefore, this potential hydrogen bond is conserved in the interaction of the *cis* conformer of LSTa with HA.

The Sia moiety of both *trans* and *cis* conformers of LSTa forms, via its carboxylic O1A oxygen, a potential hydrogen bond with the amino (nitrogen ND2) group of the amide of Asn137 side chain of L226Q HA (Figures 35 and 36; Figure 40). The 4-hydroxylic group of Gal-2 of the *cis* conformer of LSTa also forms a potential hydrogen bond with the same amide amino (ND2) group of Asn137. This residue is conserved between L226Q HA and X-31 HA, but this interaction is not observed in the complex of X-31 HA with LSTa (*cis* configuration). Figure 39A shows the superposition of the L226Q HA-*cis* LSTa complex with the X-31 HA-LSTa complex. The side chain of Asn137 in the X-31 HA–LSTa complex is positioned away from Gal-2, precluding the formation of the potential hydrogen bond of the amide nitrogen (ND2) with the 4-OH group of Gal-2 seen in the L226Q HA–*cis* LSTa complex. In H9 swine HA, which recognises the *cis* configuration of the LSTa α2-3 linkage, the short side chain of Ser137 and the axial positioning of the 4-OH group of Gal-2 in LSTa do not favour the formation of a hydrogen bond (Figures 11A and 39B). Therefore, the potential interactions of the side chain of Asn137 with the 4-OH of Gal-2 and the carboxylic group of Sia, and the presence of the polar side chain of Gln226, with the potential to hydrogen bond with the carboxylic and 8-OH groups of Sia (Figure 40), may be important for the flexibility of the L226Q HA in recognising two different conformers of LSTa.

4.7.2 Structure of L226Q HA in complex with LSTc

The electron density on the RBS of L226Q HA mainly represents the Sia moiety of LSTc (Figures 33 and 41). It was not possible to model the Sia in one of RBS of L226Q HA. A similar crystal contact to the one observed in the complex of LSTa with L226Q HA seems to affect the binding of LSTc to that particular site (see *4.6.3*). The modelling of only the Sia moiety of LSTc on the HA RBS was also observed in the structure of H5 avian HA in complex with LSTc (Figure 9B) (Ha et al., 2001).

4.7.2.1 Atomic interactions between LSTc and L226Q HA

A schematic representation of the atomic interactions between the Sia residue of LSTc and L226Q HA is shown in Figure 42. For comparison, a list of the interatomic distances of the Sia moiety of LSTc with L226Q HA and the other HAs whose structure in complex with LSTc has been already determined is shown in Table 11, Appendix IV. These atomic interactions correspond to potential hydrogen bonds between receptor analogue and protein.

The pattern of putative hydrogen bonds of the Sia moiety of LSTc with the RBS of L226Q HA (Figure 41) is very similar to that observed in the so-far determined structures of LSTc and LSTa in complex with HAs of different hosts and subtypes (Figures 7-13; Table 11, Appendix IV). As previously mentioned (see *1.6.7*, and *4.7.1.2*), the pattern of interactions of the Sia of different analogues with the RBS of HA is generally conserved. In contrast with HAs having Leu226, which is only involved in van der Waals interactions (Figures 7A, 7B, 11A, 11B), in L226Q HA the side chain of Gln226 forms a pair of hydrogen bonds with the Sia of either LSTc or LSTa also observed in other HAs that have a Gln at position 226, namely H3 avian HA, H5 avian HA, H7 avian HA, H1 human HA and H1 swine HA (Figures 40 and 42; Tables 10 and 11, Appendix IV). These conserved hydrogen

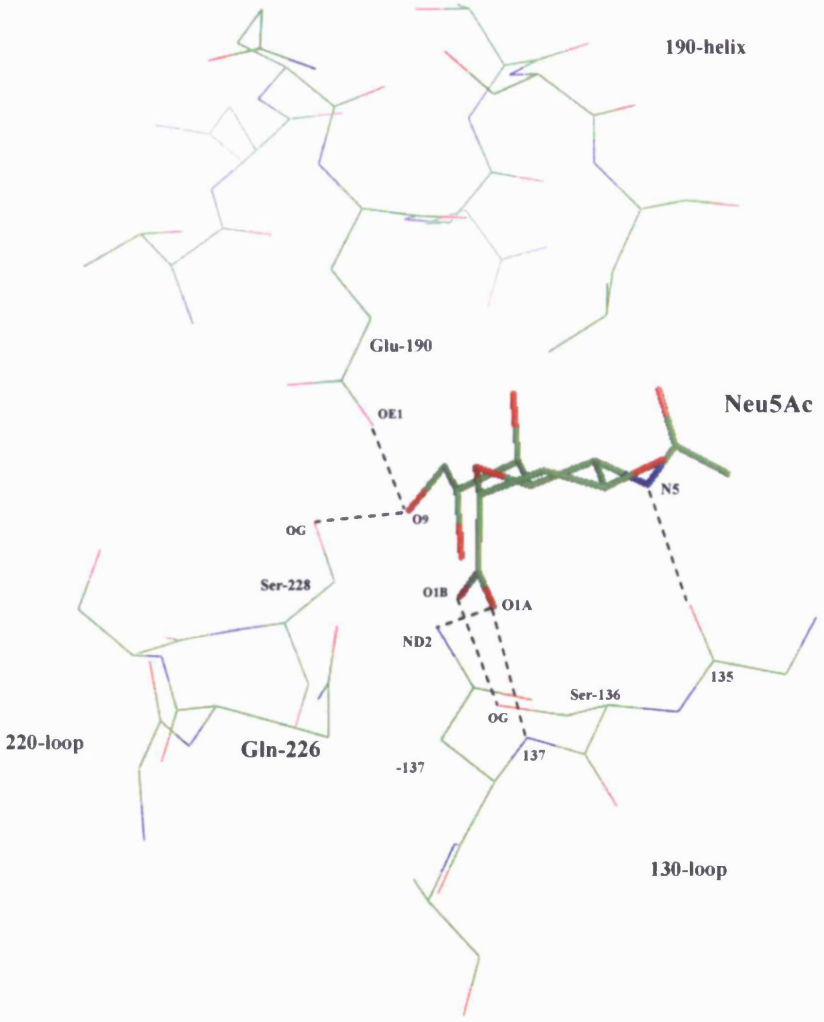


Figure 41. Binding of LSTc to L226Q HA

Only the Sia of LSTc is observed in two of the RBSs of L226Q HA. Selected potential hydrogen bonds between analogue and L226Q HA are shown as dashed lines. Atoms are coloured as: green for carbon, blue for nitrogen, and red for oxygen. Relevant residues of the HA RBS and atoms of Sia and the HA RBS, namely those involved in the selected interatomic contacts, are specifically identified (see: Figure 42; *4.7.2.1* and *4.8*; Table 11, Appendix IV). The three main secondary structural elements of the HA RBS - 130 loop, 190 helix and 220 loop - are also indicated.

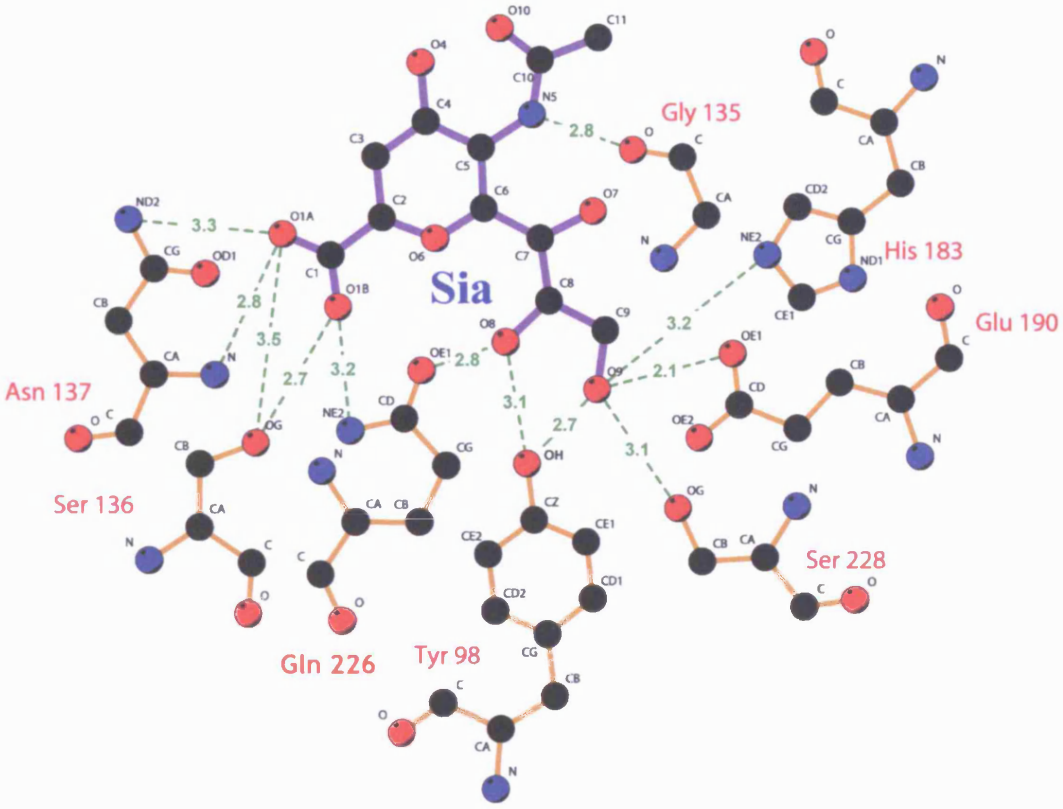


Figure 42. Schematic diagram of the interactions of LSTc with the residues of the RBS of L226Q HA

Potential hydrogen bonds (shown as dashed lines) between LSTc and the residues of the RBS of L226Q HA are shown. Only the Sia of LSTc is shown, since this was the only residue of the ligand that could be fitted into the electron density (see *4.6*). The interatomic interaction distances corresponding to the represented hydrogen bonds are also indicated (see, also, Table 11, Appendix IV). Atoms (shown as balls) are coloured as: black for carbon, blue for nitrogen, and red for oxygen. Covalent bonds (shown as sticks) in LSTa and the L226Q HA residues are coloured in purple and brown, respectively. The identity of main-chain and side chain atoms is indicated.

The figure was generated using the program LIGPLOT (Wallace et al., 1995).

bonds of the side chain of Gln226 are formed between the amide nitrogen and the carboxylic (O1B) group of the Sia of LSTc, and between the amide carbonyl and the 8-hydroxyl group of the Sia of LSTc. The L226Q HA, as the X-31 HA, forms a direct hydrogen bond, via the side chain hydroxyl of Ser228, with the 9-hydroxylic group of the glycerol side chain of the Sia residue of LSTc. This interaction is conserved in the complexes of L226Q HA with LSTc and with both the *cis* and *trans* conformers of LSTa (see, above, *4.7.1.2;* Figures 40 and 42). Also, in L226Q HA-LSTc, as in the complex of X-31 HA with LSTa, the amide nitrogen (ND2) of Asn137 forms a hydrogen bond with the carboxylic (O1A) group of the Sia of LSTc (Figure 42; Tables 10 and 11, Appendix IV).

4.8 Discussion

4.8.1 Structure of L226Q HA in complex with LSTa

In this study, the X-ray crystal structure of the LSTa-L226Q HA complex revealed that the L226Q HA binds both the *trans* and *cis* configurations of the α2-3 linkage of LSTa, each on a different RBS (Figures 35 and 36). This is the first description of two different conformers of LSTa being able to simultaneously bind to HA, and also the first description of the binding of the LSTa *cis* conformer to an HA having Gln at position 226 of the RBS.

4.8.1.1 Configurations of LSTa in complex with different HAs

The crystal structures of HA-LSTa complexes determined so far have shown that the HAs of different hosts and subtypes bind only one configuration of LSTa. H3 avian HA, H5 avian HA and H7 avian HA, which show binding preference for α2-3-linked sialosides (Matrosovich et al., 1993; 1997; 1999; Gambaryan et al., 1997), and H1 human HA, which shows dual binding specificity for α2-3-linked and α2-6-linked sialosides (Rogers and Paulson, 1983; Rogers and D'Souza, 1989), revealed in their crystal structures in complex with LSTa (Ha et al., 2003; Ha et al., 2001; Russell et al., 2006; Gamblin et al., 2004) that they only recognised the *trans* conformer of the avian receptor analogue (Figures 8A, 9A, 10, and 12A). Therefore, it seems that the LSTa conformer adopting the *trans* configuration around the glycosidic oxygen of the α2-3 linkage has the energetically most favourable conformation to bind the RBS of H3 avian HA, H5 avian HA, H7 avian HA, and H1 human HA. On the other hand, X-31 HA and H9 swine HA prefer binding to α2-6-linked sialosides (Nobusawa et al., 1991; reviewed in Skehel et al., 2000) and in the crystal structures of the complexes LSTa-X-31 HA and LSTa-H9 swine HA (Eisen et al., 1997; Ha et al., 2001), only the *cis* conformer of LSTa binds the RBS of both these HAs (Figures 7A and 11A). This suggests that the LSTa conformer having the *cis* configuration

of the α2-3 linkage has the energetically most favourable conformation to fit in the RBSs of X-31 HA and H9 swine HA. Therefore, HAs showing a strong binding preference for α2-3-linked sialosides, which correlates with the general receptor preference of avian viruses, or a dual binding specificity for α2-3 and α2-6 linkages (like the H1 human HA), seem to prefer the *trans* configuration of α2-3 linkages, and HAs with a strong binding preference for α2-6-linked sialosides, which correlates with the general receptor preference of human viruses, presumably bind better to the *cis* configuration of α2-3 linkages.

Here, in contrast with the specific binding of just one configuration of the α2-3 linkage by all the available HA RBSs, as seen in all structures of other HAs in complex with LSTa, each of the two available RBSs of a L226Q HA trimer specifically binds to a different configuration of the α2-3 linkage of LSTa. The L226Q HA RBS1 binds the *trans* configuration of LSTa and RBS2 binds the *cis* configuration (Figures 35 and 36). This seems to be mainly due to the positional plasticity of the side chain of Asn137 of the L226Q HA RBS (see *4.7.1.1* and Figure 37). The *trans* and *cis* configurations of the α2-3 glycosidic linkage of the extended conformation of LSTa correspond to the low-energy conformers *anti* and *syn*, respectively, of α2-3-linked sialosides identified in modelling studies, and both conformers were considered to form significant populations in solution (Breg et al., 1989; Poppe et al., 1989; Sabesan et al., 1991; Eisen et al., 1997; Ha et al., 2001). Therefore, the *trans* and *cis* LSTa conformers in solution may be equally available for binding to a HA RBS. The fact that Sia and Gal are the two visible saccharide residues of LSTa on both available sites of L226Q HA and the comparable quality of the electron density for the two LSTa conformers (see *4.6.2*; Figure 32), suggest that the binding affinity of RBS1 for the *trans* configuration of LSTa is similar to that of RBS2 for the *cis* configuration of LSTa. Both the *trans* and *cis* configurations of α2-3 linkages have been

observed in other protein-sialoside complexes (reviewed in Imberty and Pérez, 2000).

4.8.1.2 Comparison with H3 avian HA and X-31 HA in complex with LSTa

The L226Q HA is a single amino acid mutant of the X-31 HA, which corresponds to the HA of the 1968 HK pandemic virus. The H3 avian HA of the duck/Ukraine/1/63 virus, which is a potential precursor of the 1968 HK pandemic virus, shows only twenty amino acids differences to the X-31 HA and is structurally highly homologous to the L226Q HA (Ha et al., 2001). These three HAs of the H3 subtype have different pairs of amino acids at positions 226 and 228, which have shown to be determinants of receptor binding specificity in HAs of different subtypes (Connor et al., 1994): the X-31 HA has Leu226/Ser228 and binds preferably to α2-6 linkages, the L226Q HA has Gln226/Ser228 and prefers α2-3 linkages, and the H3 avian HA has Gln226/Gly228 and a binding preference for α2-3 linkages. The comparison of the crystal structures of L226Q HA in complex with LSTa and LSTc with those of the receptor analogues in complex with X-31 HA and H3 avian HA is relevant for the understanding of the molecular mechanisms underlying the conversion of a human virus HA into an avian virus HA (see below).

The high structural similarity of the RBS of L226Q HA to that of the H3 avian HA (Ha et al., 2003) might imply a binding preference of the L226Q HA for the *trans* configuration of LSTa. The intermediate mode of binding of L226Q HA to the LSTa conformers in the crystal may have been caused by particular experimental circumstances. The specific crystal contacts between symmetry-related molecules in L226Q HA might have caused slight adjustments in the L226Q HA that favoured the stable binding of two different configurations of LSTa (see *4.6.3*). The crystal contact affecting the L226Q HA_1 subunit containing the RBS2 (Figure 34), which binds the *cis* conformer of LSTa, may induce

slight changes in the positioning of specific residues that might have created the conditions to stabilise the binding of the *cis* configuration of the α2-3 linkage (see *4.6.3.1*).

In comparison with the structure adopted when bound to the H3 avian HA (Dk/Ukraine/1/63) the Sia and Gal residues of LSTa are shifted to the front of the site and upwards in the RBS1 of L226Q HA (Figure 38A). This positional shift of the receptor analogue has had the major consequence of positioning the glycosidic oxygen away from the side chain of Gln226 (~ 2 Å upward shift in comparison with H3 avian HA), precluding the formation of a stabilising hydrogen bond with the amide nitrogen. Nonetheless, the 4-OH of Gal-2 is still able to hydrogen bond to the amide carbonyl of the side chain of Gln226 (see *4.7.1.2*). Therefore, in contrast with the H3 avian HA (Ha et al., 2003), the L226Q HA is not able to form the linkage recognition motif specific for the *trans* configuration of the α2-3 linkage, as first described for the H5 avian HA in complex with LSTa (Ha et al., 2001) (see *4.7.1.2*). This indicates a lower affinity of the L226Q HA, relatively to the H3 avian HA, for the *trans* conformer of LSTa, since the potential for stabilisation of the asialo part of LSTa in the *trans* configuration in the L226Q HA is weak, which might be reflected in the high *B* values and poor definition of the electron density. The binding of L226Q HA to LSTa also includes the interaction with the *cis* conformer, but the just described possible less stable binding of L226Q HA to the *trans* conformer in comparison with the H3 avian HA, might correlate with the significantly higher binding affinity for 3'SL showed by the H3 avian HA, in competitive binding assays of solid phase immobilised virus (Matrosovich et al., 1993). This may also be connected with the fact that only the first two residues of LSTa are seen in the structure of the L226Q HA-LSTa complex, while the GlcNAc-3 residue is also seen in all the available structures of avian HAs in complex with LSTa (Figures 8A, 9A and 10). Of the nineteen amino acid differences between the L226Q HA and H3 avian HA only four (N137S,

S193N, S227P, and S228G) are at the RBS, and of these only the amino acids at positions 137 and 228 form interactions with the receptor analogue (Table 10, Appendix IV; Figures 40 and 8A). The L226Q HA forms, through the side chain of Ser228, a direct hydrogen bond with the glycerol 9-OH group of the Sia of LSTa, while in the H3 avian HA that hydrogen bond is water-mediated. In both HAs, the main-chain amide nitrogen of amino acid 137 hydrogen binds to the carboxylic group of Sia, but the L226Q HA is also able to form a hydrogen bond with the same carboxylic group through the side chain of Asn137. The different residues at positions 227 and 228 contribute for slight conformational differences between the RBSs of H3 avian HA and L226Q HA (see *Introduction*, *1.6.6*) (Ha et al., 2003). These minor conformational differences between sites and the two just described extra hydrogen bonds that L226Q HA forms with the Sia of the LSTa *trans* conformer might contribute for the observed differences in conformation that the *trans* LSTa conformer adopts in the complexes with the L226Q HA RBS1 and H3 avian HA (Figure 38A).

The superposition of the RBS2 of L226Q HA and X-31 HA in complex with the *cis* conformer of LSTa shows that the positioning of the Sia moiety of the receptor analogue is very similar in both structures and that the Gal-2 is shifted upwards on L226Q HA in comparison with X-31 (Figure 39A). However, in spite of a higher positioning on the site, the 4-OH and 6-OH groups of Gal-2 can still form hydrogen bonds with the L226Q HA (Figure 40B). Nonetheless, the unique ability of Asn137 of the RBS2 of L226Q HA to form a potential hydrogen bond with the 4-hydroxylic group of Gal-2 of the LSTa *cis* conformer may be essential to stabilise the binding of the *cis* configuration of the α2-3 linkage. This way the LSTa *cis* conformer may adopt a bound energy state low enough to stably bind the RBS2 of L226Q HA, as discussed in *4.6.1.2*. The unique pair of residues

Gln226 and Asn137 with hydrogen bond potential in the L226Q HA (the X-31 HA has Leu226) may be an important factor in the flexibility of receptor configuration recognition.

4.8.1.3 Comparison with H5 avian HA, H7 avian HA, H1 human HA, and H9 swine HA in complex with LSTa

The LSTa *trans* conformer is also shifted in its binding to the RBS1 of L226Q when in comparison with that seen in the complexes with H5 avian HA, H7 avian HA and H1 human HA (Figure 38B-D). The positional shift of the glycosidic oxygen of LSTa in the *trans* LSTa-L226Q HA complex overlaid on the *trans* LSTa-H3 avian HA complex is comparable to that seen in the superposition with the *trans* LSTa-H5 avian HA and *trans* LSTa-H7 avian HA, but it is less apparent in the comparison with the H1 human HA-*trans* LSTa complex. This correlates with the longer distance between the glycosidc oxygen of LSTa and the amide nitrogen of the Gln226 side chain seen in the H1 human HA in comparison with that of the avian HAs (Tables 10, Appendix IV). The conformations of the RBSs of L226Q HA, H5 avian HA and H7 avian HA are very similar (Russell et al., 2006). Both avian HAs have a Ser at position 137, and the H7 avian HA forms a hydrogen bond with the Sia carboxylic group through the hydroxyl group of the side chain of Ser137 (Table 10, Appendix IV). The H1 human HA has a different RBS geometry (Gamblin et al., 2004). The different conformation of the L226Q HA RBS1 and the hydrogen bonds to LSTa from Ser228 and Asn137, whose side chain is differently positioned in comparison with that of Ser137 in H7 avian HA, may be contributing factors for the different orientation of the *trans* conformer of LSTa on the RBS1 of L226Q HA.

The unfavourable positioning of the glycosidic oxygen in the *cis* conformation of the α2-3 linkage over the polar side chain of Gln226, in the complex of the LSTa conformer with

the L226Q HA, seems to be compensated for by the formation of the hydrogen bond between the side chain of Asn137 and the 4-OH group of Gal-2. This interaction is not formed in the complex of *cis* LSTa with the H9 swine HA (see *4.7.1.2*, and Figure 39B), which like X-31 has Leu226 with a nonpolar side chain, but an avian-like Gly228. The L226Q HA has an avian-like Gln226, but a human-like Ser228. These amino acid differences at positions 226 and 228 may be an important factor for the differences in positioning of the LSTa *cis* conformer between L226Q HA, X-31 HA and H9 swine HA.

4.8.2 Structure of L226Q HA in complex with LSTc

In the crystal structure of the L226Q HA-LSTc complex, only the Sia moiety was ordered on the RBS of L226Q HA (Figure 41). The Sia portion of LSTc established conserved interactions with the RBS, while the asialo part of the analogue may have a high degree of freedom. This ligand disorder could have been due to a low binding affinity of LSTc to L226Q HA or the similar binding of the Sia moiety of different conformations of the analogue to the same RBS of L226Q HA, as suggested for the binding of only the Sia of LSTc to the H5 avian HA (Ha et al., 2001). H5 influenza viruses show negligible binding to α2-6-linked Sia-Gal linkages (Matrosovich, 1999), and in NMR studies of BHA binding to soluble sialosides (Sauter et al., 1989) the L226Q HA showed a very low binding affinity for 6'SL (K_D ~ 6 mM). The binding of only Sia to the RBS of H5 avian HA and L226Q HA in both the structures of the complexes H5 avian HA-LSTc and L226Q HA-LSTc is consistent with the low binding affinity of these two HAs for α2-6 linkages.

4.8.2.1 Configuration of LSTc in complex with different HAs

In the crystal structures of LSTc in complex with H3 human HA (X-31), H3 avian HA, H9 swine HA, H1 human HA and H1 swine HA (Eisen et al., 1997; Ha et al., 2003; HA et al.,

2001; Gamblin, 2004), the receptor analogue is bound in the *cis* configuration (Figures 7B, 8B, 11B, 12B, and 13B). This is in agreement with the determined predominance of α2-6 linkages in the *cis* configuration in solution, with a *cis*-to-*trans* ratio of 9:1 (Poppe et al., 1992). In the *cis* configuration, LSTc points the C6 linkage atom of Gal-2 down toward favourable nonpolar contacts with the side chain of Leu226 of X-31 HA and H9 swine HA (Figures 7B and 11B). H1 human HA and H1 swine HA have a different site geometry, which allows them to bind the LSTc conformer with high affinity, in spite of having Gln at position 226, and forming several interactions with the asialo part of the analogue (Figures 12B and 13B).

When in complex with the H3 avian HA (Figure 8B), in order to cope with the energetically unfavourable positioning of the nonpolar C6 linkage atom toward the polar side chain of Gln226, the Sia ring was shifted slightly upwards, had its orientation twisted and rotated in a way that the C6 methylene group of Gal-2 was slightly removed from the side chain of Gln226. As mentioned just above, in solution, α2-6 linkages are predominantly in the *cis* (or *syn*) configuration. Therefore the *cis* conformer is the energetically more stable conformer of LSTc in solution and is thus more available for binding to L226Q HA. The binding of the LSTc *cis* conformer to the L226Q HA would unfavourably position the C6 linkage atom toward the polar side chain of Gln226, and this would probably account for the observed lack of electronic density for the asialo part of LSTc bound to the L226Q HA RBS (Figure 33). In spite of the close structural similarities between the L226Q HA and the H3 avian HA, the latter can bind the Sia-Gal linkage of LSTc, but the L226Q HA can only bind the Sia moiety. The small differences in geometry of the unbound RBS of L226Q HA from that of the H3 avian HA (Ha et al., 2003) and possible further slight changes resulting from crystal contacts (see *4.6.3*), might have

contributed for an unstable binding of the *cis* conformer of LSTc to the L226Q HA under study.

If the L226Q HA bound LSTc in the *trans* configuration, the glycosidic oxygen of the α2-6 linkage would be favourably positioned to form a stabilising hydrogen bond with the amide nitrogen of the side chain of Gln226, but the 4-OH group of Gal-2 would be too far away from hydrogen bonding to the amide carbonyl of Gln226, to complete the formation of a motif mimicking the α2-3 linkage-recognition motif (see *4.7.1.2*). The binding of the *trans* conformer of LSTc to the L226Q HA would be very unlikely, since it is much less stable than the *cis* conformer and its potential for hydrogen bonding with the L226Q HA is weak.

4.8.3 Comparison of the binding of LSTa and LSTc to the L226Q HA

The above-described crystal structural data are consistent with a higher binding affinity of the L226Q HA to LSTa in comparison with LSTc. Only the Sia moiety of LSTc is seen bound on the RBS of L226Q HA, while both the terminal linkage residues, Sia and Gal, of LSTa could be modelled on the L226Q HA RBS and potential interactions of Gal-2 of the LSTa *cis* conformer with protein residues could be assigned. This correlates with binding data for both the X-31/HS virus and L226Q HA. Assays of influenza virus-induced haemagglutination of modified erythrocytes (Rogers et al., 1983; Anders et al., 1986; Daniels et al., 1987), competitive assays of virus binding to soluble receptor analogues (Matrosovich et al., 1993; 1997), and NMR studies of BHA binding to soluble sialosides (Sauter et al., 1989) showed that the L226Q mutation induces a shift in receptor-linkage preference of the parent X-31 virus from α2-6-terminated to α2-3-terminated receptors. The latter studies revealed very small differences in the binding affinities of X-31 HA and L226Q HA to 3'SL and 6'SL: L226Q HA binds 3'SL 2 fold stronger than 6'SL, and X-31

HA binds 6'SL 1.5 fold stronger than 3'SL; the affinity of LSTa for X-31 HA is slightly lower than that of 3'SL. However no other measurements were done using either LSTa or LSTc. The comparative binding affinities of X-31 HA and L226Q HA for LSTa and LSTc would determine if they are stronger than those of 3'SL and 6'SL.

4.8.4 Role of L226Q mutation of HA in the interspecies transmission of H3 influenza viruses

The induced shift in virus receptor binding preference by the L226Q mutation of the HA RBS corresponds to a shift in viral receptor recognition from a human-like receptor to an avian-like receptor. The X-31/HS virus is an egg-adapted virus, and in this study it has grown to high titters in embryonated hen's eggs upon inoculation of the allantoic cavity, which is rich in α2-3-linked sialosides and does not have α2-6-linked ones (Ito et al., 1997). Different binding studies of human reassortant viruses containing single-site mutant H3 HAs suggest that the single L226Q mutation is necessary but not sufficient to allow the virus to replicate in the duck intestine. The additional S228G mutation seems to be required to make the virus able to efficiently attach to the target cells in the duck intestinal mucosa (Hinshaw et al., 1983; Naeve et al., 1984; Vines et al., 1998). The substitution L226Q in the human H3 HA confers binding to the Sia(α2-3)Gal-containing receptor, and the second mutation S228G enhances the binding affinity enough to allow virus replication (Vines et al., 1998). More recently, the replication in duck intestine of the double mutant L226Q/S228G of the H3 human reassortant virus was associated with the recognition of Neu5Gc(α2-3)Gal-terminated receptors, abundant in the lower intestine, namely crypt cells (Ito et al., 2000).

Presently, there is no information from structural studies that could suggest a role for the second S228G mutation in the transformation of the human X-31 HA into an "avian-like"

HA. Nor there is structural information for a L226Q/S228G double mutant of a H3 human HA. Residue 226 is always Gln in avian HAs, and never Gln in H3 human HAs, but always a nonpolar residue, usually Leu and occasionally Val or Ile (Ha et al., 2003). Residue 228 is Gly or Ser in avian HAs but always Ser in H3 human HAs (Bean et al., 1992; Connor et al., 1994). The residual structural differences between the RBS of L226Q HA and that of H3 avian HA include other RBS residues (N137S, S193N, and S227P), which contribute to the geometry of the site (see *Introduction*, *1.6.6*) and probably to make it an "avian RBS", with the side chain of Gln226 being able to form a α2,3 linkage-recognition motif for the *trans* configuration of LSTa (Ha et al., 2001) (see *4.7.1.2*). The structure of LSTa, and a Neu5Gc(α2-3)Gal-terminated sialoside (see just above), in complex with a H3 human HA double-site L226Q/S228G mutant would probably bring new information for the molecular transformation of a human HA into an avian-like HA at the level of receptor binding. The second mutation S228G added to the L226Q HA might be sufficient for the formation of the linkage-recognition motif specific for the *trans* configuration of α2-3 linkages, as seen in the complexes of avian HAs with LSTa whose crystal structures have been determined (Ha et al., 2001, 2003; Russell et al., 2006) (see *1.6.7.1* and *4.7.1.2*).

On the other hand, the crystal structure of the L226Q/S228G double mutant of the X-31 HA in complex with LSTa and LSTc would allow insights into the molecular mechanism of the conversion of an avian virus into a human virus, with pandemic characteristics, at the level of receptor recognition. In support of this possibility, a recombinant virus harbouring the L226Q/S228G double mutant of the X-31 HA has recently shown to be able to replicate and spread in human airway epithelial cultures, even though less efficiently than the virus harbouring the X-31 HA human counterpart (Matrosovich et al., 2007).

The L226Q HA may correspond to an intermediate HA in the binding to α2-3-linked receptors between a human-like recognition, specific for the *cis* configuration of LSTa - like that seen in H3 human HA and H9 swine HA, which have a Leu at position 226 of the RBS (Eisen et al., 1997; Ha et al., 2001) – and an avian-like recognition, specific for the *trans* configuration of LSTa – like that seen in the avian HAs H3 avian HA, H5 avian HA and H7 avian HA, and the H1 human HA, all having a Gln at position 226 of the RBS (Ha et al., 2001, 2003; Russell et al., 2006; Gamblin et al., 2004).

4.9 Conclusions

The crystal structure of L226Q HA in complex with LSTa, determined in this study, indicates that the L226Q HA trimer has two RBSs with enough structural differences to give them different binding specificities for the configuration of the α2,3 linkage. It is possible that, due to lattice contacts and conformational plasticity associated with the side chain of Asn137, very small changes in the structure of one RBS in the L226Q HA trimer might have been translated into changes in specificity for the configuration of the α2,3 linkage. The preference of RBS1 for the *trans* configuration of the Sia-Gal linkage of LSTa and RBS2 for the *cis* configuration represents a new kind of behaviour in receptor linkage recognition by HA, since in all the so-far determined structures of α2,3- or α2,6-linked sialosides in complex with HAs of different hosts and subtypes only one configuration is seen.

The LSTa *trans* and *cis* conformers may be equally populated in solution, thus their binding specifically and separately to the L226Q HA RBS1 and RBS2 would suggest that the two L226Q HA RBSs have a similar binding affinity for the different LSTa conformer bound. The finding that in both RBS1 and RBS2 of L226Q HA only the Sia and Gal-2 residues of LSTa are ordered is consistent with them having similar binding affinities for the *trans* and *cis* LSTa conformer, respectively.

The L226Q HA in its binding to the *trans* conformer of LSTa does not form the α2,3 linkage specific motif seen in avian HAs, due to a positional shift of the bound ligand. In the RBS1 of the L226Q HA, the side chain of Gln226 forms a hydrogen bond with the 4-hydroxyl group of Gal-2, but not with the glycosidic oxygen of the α2,3 linkage of LSTa. This indicates a lower affinity, in comparison with avian HAs, of L226Q HA for the *trans*

LSTa. However, the side chain of Asn137 forms a hydrogen bond with the carboxylic group of Sia that might add to the stability of the LSTa in the new position. Also, the side chain of Ser228 forms a direct hydrogen bond with the 9-hydroxyl group of the glycerol moiety of Sia. Therefore, this new pattern of interactions in the RBS1 of L226Q HA might enhance the affinity of the site for the new positioning of the *trans* LSTa conformer.

The binding of the LSTa *cis* conformer to the L226Q HA RSB2 corresponds to an interaction not previously seen in the structures of other HA-*cis* LSTa complexes. The *cis* configuration of the α2-3 linkage positions the glycosidic oxygen away from a favourable interaction with the side chain of Gln226. Here, besides the usual hydrogen bond between main chain carbonyl of residue 225 and the extracyclic 6-OH of Gal-2, the side chain of Asn137 forms a hydrogen bond with the 4-OH group of Gal-2. This additional hydrogen bond may compensate for the energetically unfavourable positioning imposed by the *cis* configuration on the α2,3 linkage glycosidic oxygen.

The L226Q HA only binds the Sia moiety of LSTc, suggesting a low affinity for the human receptor analogue. This is consistent with data from binding assays indicating a preference of L226Q HA for α2-3-linked receptor analogues.

Therefore, despite its high homology with the H3 avian HA and an avian-like receptor binding preference, the L226Q HA recognises both the *trans* configuration, preferred by avian viruses, and the *cis* configuration, preferred by human viruses. This suggests that the L226Q HA represents an intermediate HA for avian receptor binding between a human HA and an avian HA. The additional mutation S228G may be necessary for the conversion

of the human X-31 HA into an avian-like HA, with a specific high affinity for the *trans* conformer of LSTa, as seen in avian HAs.

The binding affinities of HA for α2,3-linked or α2,6-linked soluble sialosides are very weak (K_Ds in the range of 1-6 mM) and this has implications for the structural characteristics and definition of the complexes. Nonetheless it has been possible to characterise crystallographically the protein and ligand structures sufficiently to identify their interactions and to define their ligand conformations. However small structural changes will not always have been detected in the x-ray analysis owing to the limited resolution of the data (2.8-2.9 Å) and the ill defined electron density.

4.10 Future Experiments

4.10.1 Improvement of the crystal structure of L226Q HA in complex with LSTa and LSTc

The limited resolution of the crystal structures of the L226Q HA-LSTa and L226Q HA-LSTc complexes determined in this study has imposed significant limitations to the determination of the interactions of the receptor analogues with the L226Q HA. The errors of the refined atomic coordinates are high, affecting the accuracy of the assigned hydrogen bonds between protein and ligand. Also, due to the fact that no water molecules were added to the model, potential hydrogen bonds between the receptor analogues and the L226Q HA might have been left unassigned.

As future work, the quality improvement of the L226Q HA crystals would contribute for a more detailed picture of the interaction of the receptor analogues with the RBS of L226Q HA.

4.10.2 Determination of the crystal structure of the L226Q/S228G HA double mutant

The L226Q/S228G HA double mutant variants of H3 human influenza isolates have shown an increased affinity for α2-3-linked receptors in comparison with the mutant viruses having the L226Q HA mutation. This enhancement in receptor binding affinity showed to be essential for the replication of the variant human viruses on the duck intestine. Similarly to these H3 viruses, the L226Q/S228G HA double mutant of the X-31 virus may, not only have a higher affinity for the receptor in comparison with the L226Q HA, but also different receptor binding characteristics.

As future work, the crystal structure determination of the L226Q/S228G X-31 mutant HA in complex with receptor analogues promises new insights into the relationship between the RBS sequence and ligand binding. Also, the structure determination of mutant HAs (in complex with receptor analogues) from avian or human H3/H2 virus isolates, like L226Q HA, L226Q/S228G HA, Q226L HA and Q226L/G228S HA would most probably reveal details of the molecular mechanism of the shift in viral receptor specificity of those mutants, which has implications for the interspecies transmission of influenza viruses.

4.10.3 Binding assays of L226Q HA and L226Q/S228G HA to receptor analogues

Generally, the receptor binding assays of influenza viruses have been carried out using derivatised erythrocytes, soluble sialosides and immobilised glycopolymers (reviewed in Matrosovich et al., 2006). The two latter assays have also been used to test the receptor binding activity of BHA molecules. The new assays using carbohydrate microarrays offer the advantage of testing a high number of different saccharides in one single experiment (Stevens et al., 2006a; 2006b). As future work, receptor binding assays of the X-31 HA,

L226Q HA and L226Q/S228G HA and the viruses containing these HAs would reveal new aspects of the binding mechanism of H3 influenza viruses to target cell receptors, not only at the level of single trimeric HA molecules but also in the context of virus particles, testing for cooperativity upon polyvalent binding (reviewed in Stevens et al., 2006b) (see *Introduction, 1.6.3*).

5. References

Abramson, J. and Iwata, S. (1999) Crystallization of membrane proteins. In: Protein Crystallisation: Techniques, Strategies and Tips. A Laboratory Manual. T.M. Bergfors, ed. IUL Press, La Jolla, California, *pp.* 197-210

Alexander, D.J. (2000) A review of avian influenza in different bird species. *Vet. Micribiol.* **74**, 3-13

Anders, E.M., Scalzo, A.A., Rogers, G.N. and White, D.O. (1986) Relationship between mitogenic activity of influenza viruses and the receptor-binding specificity of their hemagglutinin molecules. *J. Virol.* **60**, 476-482

Armstrong, R.T., Kushnir, A.S. and White, J.M. (2000) The transmembrane domain of influenza hemagglutinin exhibits a stringent length requirement to support the hemifusion to fusion transition. *J. Cell Biol.* **151**, 425-437

Baigent, S.J. and McCauley, J.W. (2003) Influenza type A in humans, mammals and birds, determinants of virus virulence, host-range and interspecies transmission. *Bioessays* **25**, 657-671

Barbey-Martin, C., Gigant, B., Bizebard, T., Calder, L.J., Wharton, S.A., Skehel, J.J. and Knossow, M. (2002) An antibody that prevents the hemagglutinin low pH fusogenic transition. *Virology* **294**, 70-74

Bean, W.J., Schell, M., Katz, J., Kawaoka, Y., Naeve, C., Gorman, O. and Webster, R.G. (1992) Evolution of the H3 influenza virus hemagglutinin from human and nonhuman hosts. *J. Virol.* **66**, 1129-1138

Beaton, A.R. and Krug, R.M. (1986) Transcription antitermination during influenza viral template RNA synthesis requires the nucleocapsid protein and the absence of a 5' capped end. *Proc. Natl. Acad. Sci. USA* **81**, 6282-6286

Bergfors, T.M. (1999) Protein Samples; Dynamic Light Scattering. In: Protein crystallization: Techniques, Strategies, and Tips. A Laboratory Manual. T.M. Bergfors, ed. IUL Press, La Jolla, California, *pp.* 19-38

Bizebard, T., Gigant, B., Rigolet, P., Rasmussen, B., Diat, O., Bosecke, P., Wharton, S.A., Skehel, J.J. and Knossow, M. (1995) Structure of influenza virus haemagglutinin complexed with a neutralizing antibody. *Nature* **376**, 92-94

Blackman, M.J., Fujioka, H., Stafford. W. H.L., Sajid, M., Clough, B., Fleck. S.L., Aikawa, M., Grainger, M. and Hackett, F. (1998) A subtilisin-like protein in secretory organelles of Plasmodium falciparum merozoites. *J. Biol. Chem.* **273**, 23398-23409

Boehringer Mannheim (1990) Detergents for Membrane Research, Catalogue

Bos, T.J., Davis, A.R. and Nayak, D.P. (1984) NH2-terminal hydrophobic region of influenza virus neuraminidase provides the signal function in translocation. *Proc. Natl. Acad. Sci. U S A* **81**, 2327-2331

Boulan, E.R. and Sabatini, D.D. (1978) Asymmetric budding of viruses in epithelial monolayers: a model system for study of epithelial polarity. *Proc. Natl. Acad. Sci. U S A* **75**, 5071-5075

Braam, J., Ulmanen, I. and Krug, R.M. (1983) Molecular model of a eukaryotic transcription complex: functions and movements of influenza P proteins during capped RNA-primed transcription. *Cell* **34**, 609-618

Brand, C.M. and Skehel, J.J. (1972) Crystalline antigen from the influenza virus envelope. *Nat. New Biol.* **238**,145-147

Breg, J., Kroon-Batenburg, L.M., Strecker, G., Montreuil, J. and Vliegenthart, J.F. (1989) Conformational analysis of the sialyl alpha(2----3/6)N-acetyllactosamine structural element occurring in glycoproteins, by two-dimensional NOE 1H-NMR spectroscopy in

combination with energy calculations by hard-sphere exo-anomeric and molecular mechanics force-field with hydrogen-bonding potential. *Eur. J. Biochem.* **178**, 727-739

Bui, M., Whittaker, G. and Helenius, A. (1996) Effect of M1 protein and low pH on nuclear transport of influenza virus ribonucleoproteins. *J. Virol.* **70**, 8391-8401

Bukrinskaya, A.G., Vorkunova, N.K., Kornilayeva, G.V., Narmanbetova, R.A. and Vorkunova, G.K. (1982) Influenza virus uncoating in infected cells and effect of rimantadine. *J. Gen. Virol.* **60**, 49-59

Bullough, P.A., Hughson, F.M., Skehel, J.J. and Wiley, D.C. (1994) Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* **371**, 37-43

Burnet, F.M. and Bull, D.R. (1943) Changes in influenza virus associated with adaptation to passage in chick embryos. *Aust. J. Exp. Biol. Med. Sci.* **21**, 55-69

Burnet, F.M. and Lind, P.E. (1957) Studies on filamentary forms of influenza virus with special reference to the use of dark-ground-microscopy. *Arch. Gesamte Virusforsch.* **7**, 413-428

Campbell, J.W. (1995) XDL_VIEW, an X-windows-based toolkit for crystallographic and other applications. *J. Appl. Cryst.* **28**, 236-242

Carroll, S.M., Higa, H.H. and Paulson, J.C. (1981) Different cell-surface receptor determinants of antigenically similar influenza virus hemagglutinins. *J. Biol. Chem.* **256**, 8357-8363

Claas, E.C.J., Osterhaus, A.D. M.E., Vanbeek, R., Dejong, J.C., Rimmelzwaan, G.F., Senne, D.A., Krauss, S., Shortridge, K.F. and Webster, R.G. (1998) Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus. *Lancet* **351**, 472-477

Chen, J., Wharton, S.A., Weissenhorn, W., Calder, L.J., Hughson, F.M., Skehel, J.J. and Wiley, D.C. (1995) A soluble domain of the membrane-anchoring chain of influenza virus

hemagglutinin (HA2) folds in Escherichia coli into the low-pH-induced conformation. *Proc. Natl. Acad. Sci. USA* **92**, 12205-12209

Chen, J., Lee, K.H., Steinhauer D.A., Stevens, D.J., Skehel, J.J. and Wiley, D.C. (1998a) Structure of the hemagglutinin precursor cleavage site, a determinant of influenza pathogenicity and the origin of the labile conformation. *Cell* **95**, 409-417

Chen, J., Skehel, J.J. and Wiley, D.C. (1998b) A polar octapeptide fused to the N-terminal fusion peptide solubilizes the influenza virus HA2 subunit ectodomain. *Biochemistry* **37**, 13643-13649

Chen, J., Skehel, J.J. and Wiley, D.C. (1999) N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA(2) subunit to form an N cap that terminates the triple-stranded coiled coil. *Proc. Natl. Acad. Sci. USA* **96**, 8967-8972

Chiu, M.L., Nollert, P., Loewen, M.C., Belrhali, H., Pebay-Peyroula, E., Rosenbusch, J.P. and Landau, E.M. (2000) Crystallization in cubo: general applicability to membrane proteins. *Acta Cryst.* **D56**, 781-784

Chizhmakov, I.V., Geraghty, F.M., Ogden, D.C., Hayhurst, A., Antoniou, M. and Hay, A.J. (1996) Selective proton permeability and pH regulation of the influenza virus M2 channel expressed in mouse erythroleukaemia cells. *J. Physiol.* **494**, 329-336

Choppin, P.W., Murphy, J.S. and Tamm, I. (1960) Studies of two kinds of virus particles which comprise influenza A2 virus strains. III. Morphological characteristics: independence to morphological and functional traits. *J. Exp. Med.* **112**, 945-952

Chu, C.M., Dawson, I.M. and Elford, W.J. (1949) Filamentous forms associated with newly isolated influenza virus. *Lancet* **i**, 602-603

Collaborative Computational Project, Number 4 (1994) The CCP4 Suite: Programs for Protein Crystallography. *Acta Cryst.* **D50**, 760-763

Compans, R.W., Content, J. and Duesberg, P.H. (1972) Structure of the ribonucleoprotein of influenza virus. *J. Virol.* **10**, 795-800

Compans, R.W. (1973a) Influenza virus proteins. II. Association with components of the cytoplasm. *Virology* **51**, 56-70

Compans, R.W. (1973b) Distinct carbohydrate components of influenza virus glycoproteins in smooth and rough cytoplasmic membranes. *Virology* **55**, 541-545

Connor, R.J., Kawaoka, Y., Webster, R.G. and Paulson, J.C. (1994) Receptor specificity in human, avian, and equine H2 and H3 influenza virus isolates. *Virology* **205**, 17-23

Copeland, C.S., Zimmer, K.P., Wagner, K.R., Healey, G.A., Mellman, I. and Helenius, A. (1988) Folding, trimerization, and transport are sequential events in the biogenesis of influenza virus hemagglutinin. *Cell*, **53**, 197-209

Cox, N.J. and Subbarao, K. (2000) Global epidemiology of influenza, past and present. *Annu. Rev. Med.* **51**, 407-421

Cross, K.J., Burleigh, L.M. and Steinhauer, D.A. (2001) Mechanisms of cell entry by influenza virus. *Exp. Rev. Mol. Med.* **3**, 1-18

Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y. and McPherson, A. (1994) Screening and optimization strategies for macromolecular crystal growth. *Acta Cryst.* **D50**, 414-23

Daniels, R.S., Dowie, J.C., Hay, A.J., Knossow, M., Skehel, J.J., Wang, M.L. and Wiley, D.C. (1985) Fusion mutants of the influenza virus hemagglutinin glycoprotein. *Cell* **40**, 431-439

Daniels, P.S., Jeffries, S., Yates, P., Schild, G.C., Rogers, G.N., Paulson, J.C., Wharton, S.A., Douglas, A.R., Skehel, J.J. and Wiley, D.C. (1987) The receptor-binding and membrane-fusion properties of influenza virus variants selected using anti-haemagglutinin monoclonal antibodies. *EMBO J.* **6**, 1459-1465

Dauter, Z., Dauter, M. and Rajashankar, K.R. (2000) Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. *Acta Cryst.* **D56**, 232-237

DeLano, W.L. (DeLano Scientific, San Carlos, California, U.S.A., 2006)

Doms, R.W. and Helenius, A. (1986) Quaternary structure of influenza virus hemagglutinin after acid treatment. *J. Virol.* **60**, 833-839

Drenth, J. (1999) Crystallizing a protein. In: Principles of protein X-ray crystallography. Springer-Verlag New York Inc.; 2nd ed, *pp.* 1-21

Durrer, P., Galli, C., Hoenke, S., Corti, C., Gluck, R., Vorherr, T. and Brunner, J. (1996) H+-induced membrane insertion of influenza virus hemagglutinin involves the HA2 amino-terminal fusion peptide but not the coiled coil region. *J. Biol. Chem.* **271**, 13417-13421

Eisen, M.B., Sabesan, S., Skehel, J.J. and Wiley, D.C. (1997) Binding of the influenza A virus to cell-surface receptors: structures of five hemagglutinin-sialyloligosaccharide complexes determined by X-ray crystallography. *Virology* **232**, 19-31

Fazekas de St. Groth, S. (1948) Viropexis, the mechanism of influenza virus infection. *Nature* **162**, 294-295

Fouchier, R.A.M., Munster, V., Wallensten, A., Bestebroer, T.M., Herfst, S., Smith, D., Rimmelzwaan, G.F., Olsen, B. and Osterhaus, A.D. (2005) Characterization of a novel Influenza A virus hemagglutinin subtype (H16) obtained from black-headed gulls. *J. Virol.* **79**, 2814-2822

Gallagher, P.J., Henneberry, J.M., Sambrook, J.F. and Gething, M.J. (1992) Glycosylation requirements for intracellular transport and function of the hemagglutinin of influenza virus. *J. Virol.* **66**, 7136-7145

Gambaryan, A.S., Piskarev, V.E., Yamskov, I.A., Sakharov, A.M., Tuzikov, A.B., Bovin, N.V., Nifant'ev, N.E. and Matrosovich, M.N. (1995). Human influenza virus recognition of sialyloligosaccharides. *FEBS Lett.* **366**, 57-60

Gambaryan, A.S., Tuzikov, A.B., Piskarev, V.E., Yamnikova, S.S., Lvov, D.K., Robertson, J.S., Bovin, N.V. and Matrosovich, M.N. (1997) Specification of receptor-binding phenotypes of influenza virus isolates from different hosts using synthetic sialylglycopolymers: non-egg-adapted human H1 and H3 influenza A and influenza B viruses share a common high binding affinity for 6'-sialyl(N-acetyllactosamine). *Virology* **232**, 345-350

Gambaryan, A.S., Robertson, J.S. and Matrosovich, M.N. (1999) Effects of egg-adaptation on the receptor-binding properties of human influenza A and B viruses. *Virology* **258**, 232-239

Gambaryan, A.S., Webster, R. and Matrosovich, M. (2002) Differences between influenza virus receptors on target cells of duck and chicken. *Arch. Virol.* **147**, 1197-1208

Gambaryan, A., Yamnikova, S., Lvov, D., Tuzikov, A., Chinarev, A., Pazynina, G., Webster, R., Matrosovich, M. and Bovin, N. (2005) Receptor specificity off influenza viruses from birds and mammals: new data on involvement of the inner fragments of the carbohydrate chain. *Virology* **334**, 276-283

Gamblin, S.J., Haire, L.F., Russell, R.J., Stevens, D.J., Xiao, B., Ha, Y., Vasisht, N., Steinhauer, D.A., Daniels, R.S., Elliot, A., Wiley, D.C. and Skehel, J.J. (2004) The structure and receptor binding properties of the 1918 influenza hemagglutinin. *Science* **303**, 1838-1842

Garavito, R.M. (1991) In: Crystallization of Membrane Proteins. H. Michel, ed. CRC Press, Boca Raton, Florida, *pp.* 89-105

Gething, M.J., McCammon, K. and Sambrook, J. (1986) Expression of wild-type and mutant forms of influenza hemagglutinin: the role of folding in intracellular transport. *Cell* **46**, 939-950

Glick, G.D., Toogood, P.L., Wiley, D.C., Skehel, J.J. and Knowles, J.R. (1991) Ligand recognition by influenza virus. The binding of bivalent sialosides. *J. Biol. Chem.* **266**, 23660-23669

Godley, L., Pfeifer, J., Steinhauer, D., Ely, B., Shaw, G., Kaufmann, R., Suchanek, E., Pabo, C., Skehel, J.J., Wiley, D.C. and Wharton, S.A. (1992) Introduction of intersubunit disulfide bonds in the membrane-distal region of the influenza hemagglutinin abolishes membrane fusion activity. *Cell* **68**, 635-645

Gottschalk, A. (1959) In: The Viruses. F.V.M. Burnet and W.V.M. Stanley, eds. Academic Press, New York, *pp.* 51-61

Govorkova, E.A., Murti, G., Meignier, B., de Taisne, C. and Webster, R.G. (1996) African green monkey kidney (Vero) cells provide an alternative host cell system for influenza A and B viruses. *J. Virol.* **70**, 5519-5524

Gubareva, L.V., Wood, J.M., Meyer, W.J., Katz, J.M., Robertson, J.S., Major, D. and Webster, R.G. (1994) Condominant mixtures of viruses in reference strains of influenza virus due to host cell variation. *Virology* **199**, 89-97

Gubareva, L.V., Kaiser, L., Matrosovich, M.N., Soo-Hoo, Y. and Hayden, F.G. (2001) Selection of influenza virus mutants in experimentally infected volunteers treated with oseltamivir. *J. Infect. Dis.* **183**, 523-531

Gunther, I., Glatthaar, B., Doller, G. and Garten, W. (1993) A H1 hemagglutinin of a human influenza A virus with a carbohydrate-modulated RBS and an unusual cleavage site. *Virus Res.* **27**, 147-160

Guo, Y.J., Jin, F.G., Wang, P., Wang, M. and Zhu, J.M. (1983) Isolation of influenza C virus from pigs and experimental infection of pigs with influenza C virus. *J. Gen. Virol.* **64**, 177-182

Ha, Y., Stevens, D.J., Skehel, J.J. and Wiley, D.C. (2001) X-ray structures of H5 avian and H9 swine influenza virus hemagglutinins bound to avian and human receptor analogs. *Proc. Natl. Acad. Sci. USA* **98**, 11181-11186

Ha, Y., Stevens, D.J., Skehel, J.J. and Wiley, D.C. (2002) H5 avian and H9 swine influenza virus haemagglutinin structures: possible origin of influenza subtypes. *EMBO J.* **21**, 865-875

Ha, Y., Stevens, D.J., Skehel, J.J. and Wiley, D.C. (2003) X-ray structure of the hemagglutinin of a potential H3 avian progenitor of the 1968 Hong Kong pandemic influenza virus. *Virology* **309**, 209-218

Hagen, M., Chung, T.D., Butcher, J.A. and Krystal, M. (1994) Recombinant influenza virus polymerase: requirement of both 5' and 3' viral ends for endonuclease activity. *J. Virol.* **68**, 1509-1515

Han, X., Bushweller, J.H., Cafiso, D.S. and Tamm, L. K. (2001) Membrane structure and fusion-triggering conformational change of the fusion domain from influenza hemagglutinin. *Nat. Struct. Biol.* **8**, 715-720

Hanaoka, K., Pritchett, T.J., Takasaki, S., Kochibe, N., Sabesan, S., Paulson, J.C. and Kobata, A. (1989) 4-O-acetyl-N-acetylneuraminic acid in the N-linked carbohydrate structures of equine and guinea pig alpha 2-macroglobulins, potent inhibitors of influenza virus infection. *J. Biol. Chem.* **264**, 9842-9849

Hanson, J.E., Sauter, N.K., Skehel, J.J. and Wiley, D.C. (1992) Proton nuclear magnetic resonance studies of the binding of sialosides to intact influenza virus. *Virology* **189**, 525-523

Hardy, C.T., Young, S.A., Webster, R.G., Naeve, C.W. and Owens, R.J. (1995) Egg fluids and cells of the chorioallantoic membrane of embryonated chicken eggs can select different variants of influenza A (H3N2) viruses. *Virology* **211**, 302-306

Hay, A.J. (1974) Studies on the formation of the influenza virus envelope. *Virology* **60**, 398-418

Hay, A.J., Lomniczi, B., Bellamy, A.R. and Skehel, J.J. (1977) Transcription of the influenza virus genome. *Virology* **83**, 337-355

Hebert, D.N., Foellmer, B. and Helenius, A. (1995) Glucose trimming and reglucosylation determine glycoprotein association with calnexin in the endoplasmic reticulum. *Cell* **81**, 425-433

Herrler, G. and Klenk, H.D. (1987) The surface receptor is a major determinant of the cell tropism of influenza C virus. *Virology* **159**, 102-108

Herrler, G., Hausmann, J. and Klenk, H.D. (1995) Sialic acid as receptor determinant of ortho- and paramyxoviruses. In: Biology of the sialic acids. A. Rosenberg, ed. Plenum, New York, *pp.* 215-336

Higa, H.H., Rogers, G.N. and Paulson, J.C. (1985) Influenza virus hemagglutinins differentiate between receptor determinants bearing N-acetyl-, N-glycollyl-, and N, O-diacetylneuraminic acids. *Virology* **144**, 279-282

Hinshaw, V.S., Webster, R.G., Easterday, B.C. and Bean, W.J., Jr. (1981) Replication of avian influenza A viruses in mammals. *Infect. Immun.* **34**, 354-361

Hinshaw, V.S., Webster, R.G., Naeve, C.W. and Murphy, B.R. (1983) Altered tissue tropism of human-avian reassortant influenza viruses. *Virology* **128**, 260-263

Hirst, G.K. (1941) The agglutination of red cells by allantoic fluid of chick embryos infected with influenza virus. *Science* **94**, 22-23

Honda, A., Mizumoto, K. and Ishihama, A. (2002) Minimum molecular architectures for transcription and replication of the influenza virus. *Proc. Natl. Acad. Sci. U S A*, **99** 13166-13171

Horimoto, T. and Kawaoka, Y. (2001) Pandemic threat posed by avian influenza A viruses. *Clin. Microbiol. Rev.* **14**, 129-149

Hughey, P.G., Compans, R.W., Zebedee, S.L. and Lamb, R.A. (1992) Expression of the influenza A virus M2 protein is restricted to apical surfaces of polarized epithelial cells. *J. Virol.* **66**, 5542-5552

Hughson, F.M. (1997) Enveloped viruses: a common mode of membrane fusion? *Curr. Biol.* **7**, R565-569

Hull, J.D., Gilmore, R. and Lamb, R.A. (1988) Integration of a small integral membrane protein, M2, of influenza virus into the endoplasmic reticulum: analysis of the internal signal-anchor domain of a protein with an ectoplasmic NH2 terminus. *J. Cell Biol.* **106**, 1489-1498

Hunte, C. and Michel, H. (2002) Crystallisation of membrane proteins mediated by antibody fragments. *Curr. Opin. Struct. Biol.* **12**, 503-508

Imberty, A. and Pérez, S. (2000) Structure, conformation, and dynamics of bioactive oligosaccharides: theoretical approaches and experimental validations. *Chem. Rev.* **100**, 4567-4588

Ito, T., Suzuki, Y., Takada, A., Kawamoto, A., Otsuki, K., Masuda, H., Yamada, M., Suzuki, T., Kida, H. and Kawaoka, Y. (1997) Differences in sialic acid-galactose linkages in the chicken egg amnion and allantois influence human influenza virus receptor specificity and variant selection. *J. Virol.* **71**, 3357-3362

Ito, T., Couceiro, J.N., Kelm, S., Baum, L.G., Krauss, S., Castrucci, M.R., Donatelli, I., Kida, H., Paulson, J.C., Webster, R.G. and Kawaoka, Y. (1998) Molecular basis for the generation in pigs of influenza A viruses with pandemic potential. *J. Virol.* **72**, 7367-7373

Ito, T., Suzuki, T., Takada, A., Horimoto, T., Wells, K., Kida, H., Otsuki, K., Kiso, M., Ishida, H. and Kawaoka, Y. (2000) Recognition of N-glycolylneuraminic acid linked to

galactose by the alpha2,3 linkage is associated with intestinal replication of influenza A virus in ducks. *J. Virol.* **74**, 9300-9305

Iwata, S. (2003) Methods and Results in Crystallization of Membrane Proteins. S. Iwata, ed. International University Line, *pp.* 284-297

Jancarik, J. and Kim, S.H. (1991) Sparse Matrix Sampling: a screening method for crystallization of proteins. *J. Appl. Cryst.* **24**, 409-411

Jeffery, P.K. and Li, D. (1997) Airway mucosa, secretory cells, mucus and mucin genes. *Eur. Respir. J.* **10**,1655-1662

Jennings, P.A., Finch, J.T., Winter, G. and Robertson, J.S. (1983) Does the higher order structure of the influenza virus ribonucleoprotein guide sequence rearrangements in influenza viral RNA? *Cell* **34**, 619-627

Jin, H., Subbarao, K., Bagai, S., Leser, G.P., Murphy, B.R. and Lamb, R.A. (1996) Palmitylation of the influenza virus hemagglutinin (H3) is not essential for virus assembly or infectivity. *J. Virol.* **70**, 1406-1414

Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst.* **A47**, 110-119

Karas, M., Bachmann, D. and Hillenkamp, F. (1985) Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. *Anal. Chem.* **57**, 2935-2939

Karlsson, K.–A. (1989) Animal glycosphingolipids as membrane attachment sites for bacteria. *Annu. Rev. Biochem.* **58**, 309-350

Kates, M., Allison, A.C., Tyrrell, D.A.J. and James, A.T. (1961) Lipids of influenza virus and their relation to those of the host cell. *Biochim. Biophis. Acta* **52**, 455-466

Katz, J.M., Wang, M. and Webster, R.G. (1990) Direct sequencing of the HA gene of influenza (H3N2) virus in original clinical samples reveals sequence identity with mammalian cell-grown virus. *J. Virol.* **64**, 1808-1811

Katz, J.M. and Webster, R.G. (1992) Amino acid sequence identity between the HA1 of influenza A (H3N2) viruses grown in mammalian and primary chick kidney cells. *J. Gen. Virol.* **73**, 1159-1165

Kawaoka, Y., Yamnikova, S., Chambers, T.M., Lvov, D.K. and Webster, R.G. (1990) Molecular characterization of a new hemagglutinin, subtype H14, of influenza A virus. *Virology* **179**, 759-767

Kelm, S. and Schauer, R. (1997) Sialic acids in molecular and cellular interactions. *Int. Rev. Cytol.* **175**, 137-240

Kida, H., Ito, T., Yasuda, J., Shimizu, Y., Itakura, C., Shortridge, K.F., Kawaoka, Y. and Webster, R.G. (1994) Potential for transmission of avian influenza viruses to pigs. *J. Gen. Virol.* **75**, 2183-2188

Kido, H., Yokogoshi, Y., Sakai, K., Tashiro, M., Kishino, Y., Fukutomi, A. and Katunuma, N. (1992) Isolation and characterization of a novel trypsin-like protease found in rat bronchiolar epithelial Clara cells. A possible activator of the viral fusion glycoprotein. *J. Biol. Chem.* **267**, 13573-13579

Kiessling, L.L. and Pohl, N.L. (1996) Strength in numbers, non-natural polyvalent carbohydrate derivatives. *Chem. Biol.* **3**, 71-77

Kilbourne, E.D. (1969) Future influenza vaccines and the use of genetic recombinants. *Bull. World Health Organ.* **41**, 643-645

Kilbourne, E.D. (1987) Influenza. *Plenum Pub. Corp.*, New York

Klenk, E., Fallard, H. and Lempfrid, H. (1955) The enzymatic activity of influenza virus *Hoppe-Seyler's Z. Physiol. Chem.* **301**, 235-246

Klenk, H.D. and Rott, R. (1973) Formation of influenza virus proteins. *J. Virol.* **11**, 823-831

Klenk, H.D., Wollert, W., Rott, R. and Scholtissek, C. (1974) Association of influenza virus proteins with cytoplasmic fractions. *Virology* **57**, 28-41

Klenk, H.D., Rott, R., Orlich, M. and Blodorn, J. (1975) Activation of influenza A viruses by trypsin treatment. *Virology* **68**, 426-439

Klenk, H.D., Rott, R. and Orlich, M. (1977) Further studies on the activation of influenza virus by proteolytic cleavage of the haemagglutinin. *J. Gen. Virol.* **36**, 151-161

Kleywegt, G.J. and Jones, T.A. (1996) xdlMAPMAN and xdlDATAMAN - programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets. *Acta Cryst.* **D52**, 826-828

Klimov, A.I., Bender, C.A., Hall, H.E. and Cox, N.J. (1996) Evolution of human influenza A (H2N2) viruses. In: Options for the control of influenza III. L.E. Brown, A.W. Hampson and R.G. Webster, eds. Elsevier Science B.V., *pp.* 546-552

Kobasa, D., Rodgers, M.E., Wells, K. and Kawaoka, Y. (1997) Neuraminidase hemadsorption activity, conserved in avian influenza A viruses, does not influence viral replication in ducks. *J. Virol.* **71**, 6706-6713

Kobasa, D., Takada, A., Shinya, K., Hatta, M., Halfmann, P., Theriault, S., Suzuki, H., Nishimura, H., Mitamura, K., Sugaya, N., Usui, T., Murata, T., Maeda, Y., Watanabe, S., Suresh, M., Suzuki, T., Suzuki, Y., Feldmann, H. and Kawaoka, Y. (2004) Enhanced virulence of influenza A viruses with the haemagglutinin of the 1918 pandemic virus. *Nature* **431**, 703-707

Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774—797

Kundu, A., Avalos, R.T., Sanderson, C.M. and Nayak, D.P. (1996) Transmembrane domain of influenza virus neuraminidase, a type II protein, possesses an apical sorting signal in polarized MDCK cells. *J. Virol.* **70**, 6508-6515

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685

Lamb, R.A., Zebedee, S.L. and Richardson, C.D. (1985) Influenza virus M2 protein is an integral membrane protein expressed on the infected-cell surface. *Cell* **40**, 627-633

Lamb, R.M. and Krug, M.K. (1996) Orthomyxoviridae: the viruses and their replication. In: Fields Virology. B.N. Fields, D.M. Knipe and P.M. Howley, eds. Lippincott-Raven Publishers, Philadelphia, *pp.* 1353-1395

Laskowski R A, MacArthur M W, Moss D S & Thornton J M (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, **26**, 283-291

Laver, W.G. and Valentine, R.C. (1969) Morphology of the isolated hemagglutinin and neuraminidase subunits of influenza virus. *Virology* **38**, 105-119

Lazarowitz, S.G., Compans, R.W. and Choppin, P.W. (1971) Influenza virus structural and nonstructural proteins in infected cells and their plasma membranes. *Virology* **46**, 830-843

Lazarowitz, S.G., Compans, R.W. and Choppin, P.W. (1973a) Proteolytic cleavage of the hemagglutinin polypeptide of influenza virus. Function of the uncleaved polypeptide HA. *Virology* **52**, 199-212

Lazarowitz, S.G., Goldberg, A.R. and Choppin, P.W. (1973b) Proteolytic cleavage by plasmin of the HA polypeptide of influenza virus: host cell activation of serum plasminogen. *Virology* **56**, 172-180

Lazarowitz, S.G and Choppin, P.W. (1975) Enhancement of the infectivity of influenza A and B viruses by proteolytic cleavage of the hemagglutinin polypeptide. *Virology* **68**, 440-454

Leigh, M.W., Connor, R.J., Kelm, S., Baum, L.G. and Paulson, J.C. (1995) Receptor specificity of influenza virus influences severity of illness in ferrets. *Vaccine* **13**, 1468-1473

Leslie, A.G.W., Powell, H.R., Winter, G., Svensson, O., Spruce, D., McSweeney, S., Love, D., Kinder, S., Duke, E. and Nave, C. (2002) Automation of the collection and processing of X-ray diffraction data – a generic approach. *Acta Cryst.* **D58**, 1924-1928

Levinson, B., Pepper, D. and Belyavin, G. (1969) Substituted sialic acid prosthetic groups as determinants of viral hemagglutination. *J. Virol.* **3**, 477-483

Lin, S., Naim, H.Y., Rodriguez, A.C. and Roth, M.G. (1998) Mutations in the middle of the transmembrane domain reverse the polarity of transport of the influenza virus hemagglutinin in MDCK epithelial cells. *J. Cell Biol.* **142**, 51-57

Longberg-Holm, K. and Philipson, L. (1974) Early interaction between animal viruses and cells. *Monogr. Virol.* **9**, 1-148

Martin, K. and Helenius, A. (1991) Nuclear transport of influenza virus ribonucleoproteins: the viral matrix protein (M1) promotes export and inhibits import. *Cell* **67**, 117-130

Matlin, K.S., Reggio, H., Helenius, A. and Simons, K. (1981) Infectious entry pathway of influenza virus in a canine kidney cell line. *J. Cell Biol.* **91**, 601-613

Matrosovich, M.N. (1989) Towards the development of antimicrobial drugs acting by inhibition of pathogen attachment to host cells, a need for polyvalency. *FEBS Lett.* **252**, 1-4

Matrosovich, M.N., Mochalova, L.V., Marinina V.P., Byramova, N.E. and Bovin, N.V. (1990) Synthetic polymeric sialoside inhibitors of influenza virus receptor-binding activity. *FEBS Lett.* **272**, 209-212

Matrosovich, M.N., Gambaryan, A.S. and Chumakov, M.P. (1992) Influenza viruses differ in recognition of 4-O-acetyl substitution of sialic acid receptor determinant. *Virology* **188**, 854-858

Matrosovich, M.N., Gambaryan, A.S., Tuzikov, A.B., Byramova, N.E., Mochalova, L.V., Golbraikh, A.A., Shenderovich, M.D., Finne, J. and Bovin, N.V. (1993) Probing of the receptor-binding sites of the H1 and H3 influenza A and influenza B virus hemagglutinins by synthetic and natural sialosides. *Virology* **196**, 111-121

Matrosovich, M.N., Gambaryan, A.S., Teneberg, S., Piskarev, V.E., Yamnikova, S.S., Lvov, D.K., Robertson, J.S. and Karlsson, K.A. (1997) Avian influenza A viruses differ from human viruses by recognition of sialyloligosaccharides and gangliosides and by a higher conservation of the HA receptor-binding site. *Virology* **233**, 224-234

Matrosovich, M., Gao, P. and Kawaoka, Y. (1998) Molecular mechanisms of serum resistance of human influenza H3N2 virus and their involvement in virus adaptation in a new host. *J. Virol.* **72**, 6373-6380

Matrosovich, M., Zhou, N., Kawaoka, Y. and Webster, R. (1999) The surface glycoproteins of H5 influenza viruses isolated from humans, chickens, and wild aquatic birds have distinguishable properties. *J. Virol.* **73**, 1146-1155

Matrosovich, M., Tuzikov, A., Bovin, N., Gambaryan, A., Klimov, A., Castrucci, M.R., Donatelli, I. and Kawaoka, Y. (2000) Early alterations of the receptor-binding properties of H1, H2, and H3 avian influenza virus hemagglutinins after their introduction into mammals. *J. Virol.* **74**, 8502-8512

Matrosovich, M. and Klenk, H.D. (2003) Natural and synthetic sialic acid-containing inhibitors of influenza virus receptor binding. *Rev. Med. Virol.* **13**, 85-97

Matrosovich, M.N., Matrosovich, T.Y., Gray, T., Roberts, N.A. and Klenk, H.D. (2004) Human and avian influenza viruses target different cell types in cultures of human airway epithelium. *Proc. Natl. Acad. Sci. U S A* **101**, 4620-4624

Matrosovich, M.N., Klenk, H.D. and Kawaoka, Y. (2006) Receptor specificity, host-range, and pathogenicity of influenza viruses. In, Influenza virology: current topics. Y. Kawaoka, ed. Wymondham, Norfolk: Caister Academic Press, *pp.* 95-137

Matrosovich, M., Matrosovich, T., Uhlendorff, J., Garten, W. and Klenk, H.D. (2007) Avian-virus-like receptor specificity of the hemagglutinin impedes influenza virus replication in cultures of human airway epithelium. *Virology* **361**, 384-390

McCauley, J., Bye, J., Elder, K., Gething, M.J., Skehel, J.J., Smith, A. and Waterfield, M.D. (1979) Influenza virus haemagglutinin signal sequence. *FEBS Lett.* **108**, 422-426

McDonald, I.K. and Thornton J.M. (1994) Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777-793

McGeoch, D., Fellner, P. and Newton, C. (1976) Influenza virus genome consists of eight distinct RNA species. *Proc. Natl. Acad. Sci. U S A* **73**, 3045-3049

McRee, D.E. (1999) Practical Protein Crystallography. 2nd ed. Academic Press, *pp.* 137-144; 266

Melikyan, G.B., Jin, H., Lamb, R.A. and Cohen, F.S. (1997) The role of the cytoplasmic tail region of influenza virus hemagglutinin in formation and growth of fusion pores. *Virology* **235**, 118-128

Michel, H (1990) Crystallization of membrane proteins. CRC Press, London

Millar, B.M.G. (1997) Ph. D. Thesis, University College London

Molinari, M. and Helenius, A. (2000) Chaperone selection during glycoprotein translocation into the endoplasmic reticulum. *Science* **288**, 331-333

Mosley, V.M. and Wyckoff, R.W.G. (1946) Electron micrography of the virus of influenza. *Nature* **157**, 263

Murphy, B.R. and Webster, R.G. (1996) Orthomyxoviruses. In: Fields Virology. B.N. Fields, D.M. Knipe and P.M. Howley, eds. Lippincott-Raven Publishers, Philadelphia, *pp.* 1397-1495

Murshudov, N.G., Vagin, A.A., Dodson, J.E. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst.* **D53**, 240-255

Naeve, C.W., Hinshaw, V.S. and Webster, R.G. (1984) Mutations in the hemagglutinin receptor-binding site can change the biological properties of an influenza virus. *J. Virol.* **51**, 567-569

Navaza, J. (1994) AMoRE: an automated package for molecular replacement. *Acta Cryst.* **A50**, 157-163

Navaza, J. (2001) Implementation of molecular replacement in AMoRe. *Acta Cryst.* **D57**, 1367-1372

Nermut, M.V. and Frank, H. (1971) Fine structure of influenza A2 (Singapore) as revealed by negative staining, freeze-drying and freeze-etching. *J. Gen. Virol.* **10**, 37-51

Nobusawa, E., Aoyama, T., Kato, H., Suzuki, Y., Tateno, Y. and Nakajima, K. (1991) Comparison of complete amino acid sequences and receptor-binding properties among 13 serotypes of hemagglutinins of influenza A viruses. *Virology* **182**, 475-485

Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326

Osterhaus, A.D., Rimmelzwaan, G.F., Martina, B.E., Bestebroer, T.M. and Fouchier, R.A. (2000) Influenza B Virus in Seals. *Science* **288**, 1051-1053

Oxford, J.S., Schild, G.C., Corcoran, T., Newman, R., Major, D., Robertson, J., Bootman, J., Higgins, P., al-Nakib, W. and Tyrrell, D.A. (1990) A host-cell-selected variant of influenza B virus with a single nucleotide substitution in HA affecting a potential glycosylation site was attenuated in virulence for volunteers. *Arch. Virol.* **110**, 37-46

Page, R., Grzechnik, S.K., Canaves, J.M., Spraggon, G., Kreusch, A., Kuhn, P., Raymond, C.S., Lesley, S.A. (2003) Shotgun crystallization strategy for structural genomics: an optimized two-tiered crystallization screen against the *Thermotoga maritima* proteome. *Acta Cryst.* D59, 1028-1037

Palese, P., Tobita, K., Ueda, M. and Compans, R.W. (1974) Characterization of temperature sensitive influenza virus mutants defective in neuraminidase. *Virology* **61**, 397-410

Palese P. and Shaw M.L. (2007) Orthomyxoviridae: the viruses and their replication. In: Fields Virology. B.N. Fields, D.M. Knipe and P.M. Howley, eds. Lippincott-Raven Publishers, Philadelphia, *pp.* 1647-1689

Pappin, D.J., Hojrup, P. and Bleasby, A.J. (1993) Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.* **3**, 327-332

Paulson, J.C. (1985) Interactions of animal viruses with cell surface receptors. In: The receptors. *Vol. 2.* M. Conn, ed. Academic Press, Orlando, FL., *pp.* 131-219

Paulson, J.C. and Rogers, G.N. (1987) Resialylated erythrocytes for assessment of the specificity of sialyloligosaccharide binding proteins. *Methods Enzymol.* **138**, 162-168

Perkins, D.N., Pappin, D.J., Creasy, D.M., Cottrell, J.S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567

Pinto, L.H., Holsinger, L.J. and Lamb, R.A. (1992) Influenza virus M2 protein has ion channel activity. *Cell* **69**, 517-528

Pons, M.W., Schulze, I.T., Hirst, G.K. and Hauser, R. (1969) Isolation and characterization of the ribonucleoprotein of influenza virus. *Virology* **39**, 250-259

Poppe, L., Dabrowski, J., von der Lieth, C.W., Numata, M. and Ogawa, T. (1989) Solution conformation of sialosylcerebroside (GM4) and its NeuAc (alpha 2-3)Gal beta sugar component. *Eur. J. Biochem.* **180**, 337-342

Poppe, L., Stuike-Prill, R., Meyer, B. and van Halbeek, H. (1992) The solution conformation of sialyl-alpha (2-6)-lactose studied by modern NMR techniques and Monte Carlo simulations. *J. Biomol. NMR* **2**, 109-136

Portela, A. and Digard, P. (2002) The influenza virus nucleoprotein: a multifunctional RNA-binding protein pivotal to virus replication. *J. Gen. Virol.* **83**, 723-734

Pritchett, T.J., Brossmer, R., Rose, U. and Paulson, J.C. (1987) Recognition of monovalent sialosides by influenza virus H3 hemagglutinin. *Virology* **160**, 502-506

Pritchett, T.J. and Paulson, J.C. (1989) Basis for the potent inhibition of influenza virus infection by equine and guinea pig alpha 2-macroglobulin. *J. Biol. Chem.* **264**, 9850-9858

Reid, A.H., Janczewski, T.A., Lourens, R.M., Elliot, A.J., Daniels, R.S., Berry, C.L., Oxford, J.S. and Taubenberger, J.K. (2003) 1918 influenza pandemic caused by highly conserved viruses with two receptor-binding variants. *Emerg. Infect. Dis.* **9**, 1249-1253

Reiss-Husson, F. (1992) In: Crystallisation of Nucleic Acids and Proteins, A Practical Approach. A. Ducruix and R. Giegé, eds. Oxford University Press, *pp.* 175-193

Reuter, G. and Gabius, H.J. (1996) Sialic acids structure-analysis-metabolism-occurrence-recognition. *Biol. Chem. Hoppe Seyler* **377**, 325-342

Roberts, P.C., Garten, W. and Klenk, H.D. (1993) Role of conserved glycosylation sites in maturation and transport of influenza A virus hemagglutinin. *J. Virol.* **67**, 3048-3060

Robertson, J.S., Bootman, J.S., Nicolson, C., Major, D., Robertson, E.W. and Wood, J.M. (1990) The hemagglutinin of influenza B virus present in clinical material is a single species identical to that of mammalian cell-grown virus. *Virology* **179**, 35-40

Robertson, J.S., Nicolson, C., Bootman, J.S., Major, D., Robertson, E.W. and Wood, J.M. (1991) Sequence analysis of the haemagglutinin (HA) of influenza A (H1N1) viruses present in clinical material and comparison with the HA of laboratory-derived virus. *J. Gen. Virol.* **72**, 2671-2677

Robertson, J. S. (1993) Clinical influenza virus and the embryonated hen's eggs. *Rev. Med. Virol.* **3**, 97-106

Robertson, J.S., Cook, P., Attwell, A.M. and Williams, S.P. (1995) Replicative advantage in tissue culture of egg-adapted influenza virus over tissue-culture derived virus, implications for vaccine manufacture. *Vaccine* **13**, 1583-1588

Rogers, G.N. and Paulson, J.C. (1983) Receptor determinants of human and animal influenza virus isolates: differences in receptor specificity of the H3 hemagglutinin based on species of origin. *Virology* **127**, 361-373

Rogers, G.N., Paulson, J.C., Daniels, R.S., Skehel, J.J., Wilson, I.A. and Wiley, D.C. (1983a) Single amino acid substitutions in influenza haemagglutinin change receptor binding specificity. *Nature* **304**, 76-78

Rogers, G.N., Pritchett, T.J., Lane, J.L. and Paulson, J.C. (1983b) Differential sensitivity of human, avian, and equine influenza A viruses to a glycoprotein inhibitor of infection: selection of receptor specific variants. *Virology* **131**, 394-408

Rogers, G.N., Daniels, R.S., Skehel, J.J., Wiley, D.C., Wang, X.F., Higa, H.H. and Paulson, J.C. (1985) Host-mediated selection of influenza virus receptor variants. Sialic acid-alpha 2,6Gal-specific clones of A/duck/Ukraine/1/63 revert to sialic acid-alpha 2,3Gal-specific wild type in ovo. *J. Biol. Chem.* **260**, 7362-7367

Rogers, G.N., Herrler, G., Paulson, J.C. and Klenk, H.D. (1986) Influenza C virus uses 9-O-acetyl-N-acetyneuraminic acid as a high affinity receptor determinant for attachment to cells. *J. Biol. Chem.* **261**, 5947-5951

Rogers, G.N. and D'Souza, B.L. (1989) Receptor binding properties of human and animal H1 influenza virus isolates. *Virology* **173**, 317-322

Romanova, J., Katinger, D., Ferko, B., Voglauer, R., Mochalova, L., Bovin, N., Lim, W., Katinger, H. and Egorov, A. (2003) Distinct host range of influenza H3N2 virus isolates in Vero and MDCK cells is determined by cell specific glycosylation pattern. *Virology* **307**, 90-97

Rosenberg, A., Howe, C. and Chargaff, E. (1956) Inhibition of influenza virus hemagglutination by a brain lipid fraction. *Nature* **177**, 234-235

Rossman, M.G. and Blow, D.M. (1962) The detection of sub-units within the crystallographic asymmetric unit. *Acta Cryst.* **15**, 24-31

Ruigrok, R.W., Andree, P.J., Hooft van Huysduynen, R.A. and Mellema, J.E. (1984) Characterization of three highly purified influenza virus strains by electron microscopy. *J. Gen. Virol.* **65**, 799-802

Ruigrok, R.W., Martin, S.R., Wharton, S.A., Skehel, J.J., Bayley, P.M. and Wiley, D.C. (1986) Conformational changes in the hemagglutinin of influenza virus which accompany heat-induced fusion of virus with liposomes. *Virology* **155**, 484-497

Ruigrok, R.W., Wrigley, N.G., Calder, L.J., Cusack, S., Wharton, S.A., Brown, E.B. and Skehel, J.J. (1986a) Electron microscopy of the low pH structure of influenza virus haemagglutinin. *EMBO J.* **5**, 41-49

Ruigrok, R.W., Aitken, A., Calder, L.J., Martin, S.R., Skehel, J.J., Wharton, S.A., Weis, W. and Wiley, D.C. (1988) Studies on the structure of the influenza virus haemagglutinin at the pH of membrane fusion. *J. Gen. Virol.* **69**, 2785-2795

Ruigrok, R.W.H. and Hewat, E.A. (1991) Comparison of negatively stained and frozen hydrated samples of influenza A and B and of vesicular stomatitis virus. *Micron Microsc. Acta* **22**, 423-434

Ruigrok, R.W. (1998) Structure of Influenza A, B and C viruses. In: Textbook of Influenza. K.G. Nicholson, R.G. Webster and A.J. Hay, eds. Blackwell Science, Oxford, *pp.* 29-42

Russell, R.J., Gamblin, S.J., Haire, L.F., Stevens, D.J., Xiao, B., Ha, Y. and Skehel, J.J. (2004) H1 and H7 influenza haemagglutinin structures extend a structural classification of haemagglutinin subtypes. *Virology* **325**, 287-296

Russell, R.J., Stevens, D.J., Haire, L.F., Gamblin, S.J. and Skehel, J.J. (2006) Avian and human receptor binding by hemagglutinins of influenza A viruses. *Glycoconj. J.* **23**, 85-92

Sabesan, S., Bock, K. and Paulson, J.C. (1991) Conformational analysis of sialyl-oligosaccharides. *Carbohydr. Res.* **218**, 27-54

Saito, T., Taylor, G. and Webster, R.G. (1995) Steps in maturation of influenza A virus neuraminidase. *J. Virol.* **69**, 5011-5017

Sauter, N.K., Bednarski, M.D., Wurzburg, B.A., Hanson, J.E., Whitesides, G.M., Skehel, J.J. and Wiley, D.C. (1989) Hemagglutinins from two influenza virus variants bind to sialic acid derivatives with millimolar dissociation constants: a 500-MHz proton nuclear magnetic resonance study. *Biochemistry* **28**, 8388-8396

Sauter, N.K., Hanson, J.E., Glick, G.D., Brown, J.H., Crowther, R.L., Park, S.J., Skehel, J.J. and Wiley, D.C. (1992a) Binding of influenza virus hemagglutinin to analogs of its cell-surface receptor, sialic acid: analysis by proton nuclear magnetic resonance spectroscopy and X-ray crystallography. *Biochemistry* **31**, 9609-9621

Sauter, N.K., Glick, G.D., Crowther, R.L., Park, S.J., Eisen, M.B., Skehel, J.J., Knowles, J.R. and Wiley, D.C. (1992b) Crystallographic detection of a second ligand binding site in influenza virus hemagglutinin. *Proc. Natl. Acad. Sci. USA* **89**, 324-328

Schauer, R. (1982) Chemistry, metabolism, and biological functions of sialic acids. *Adv. Carbohydr. Chem. Biochem.* **40**, 131-234

Schauer, R., Kelm, S., Schröder, C. and Miller, E. (1995) Biochemistry and role of sialic acids. In: Biology of the sialic acids. A. Rosenberg, ed. Plenum, New York, *pp.* 7-67

Schauer, R. and Kamerling, J.P. (1997) Chemistry, biochemistry and biology of sialic acids. In: Glycoproteins II (New Comprehensive Biochemistry, vol. 29b). J. Montreuil, J.F.G. Vliegenthart, and H. Schachter, eds. Elsevier Science B. V., *pp.* 243-402

Scheiffele, P., Rietveld, A., Wilk, T. and Simons, K. (1999) Influenza viruses select ordered lipid domains during budding from the plasma membrane. *J. Biol. Chem.* **274**, 2038-2044

Schild, G.C. (1972) Evidence for a new type-specific structural antigen of the influenza virus particle. *J. Gen. Virol.* **15**, 99-103

Schild, G.C., Newman, R.W., Webster, R.G., Major, D. and Hinshaw, V.S. (1980) Antigenic analysis of influenza A virus surface antigens: considerations for the nomenclature of influenza virus. Brief review. *Arch. Virol.* **63**, 171-184

Schmidt, M.F. (1982) Acylation of viral spike glycoproteins: a feature of enveloped RNA viruses. *Virology*, **116**, 327-338

Schneider, C., Newman, R.A., Sutherland, D.R., Asser, U. and Greaves, M.F. (1982) A one-step purification of membrane proteins using a high efficiency immunomatrix. *J. Biol. Chem.* **257**,10766-10769

Scholtissek, C., Burger, H., Kistner, O. and Shortridge, K.F. (1985) The nucleoprotein as a possible major factor in determining host specificity of influenza H3N2 viruses. *Virology* **147**, 287-294

Scholtissek, C., Hinshaw, V.S., and Olsen, C.W. (1998) Influenza in pigs and their role as the intermediate host. In: Textbook of influenza. K.G. Nicholson, R.G. Webster, and A.J. Hay, eds. Blackwell Science, London, *pp.* 137-145

Schuettelkopf, A.W. and van Aalten D.M.F (2004) PRODRG - a tool for high-throughput crystallography of protein-ligand complexes. *Acta Cryst.* **D60**, 1355-1363

Schulze, I.T. (1970) The structure of influenza virus. I. The polypeptides of the virion. *Virology* **42**, 890-904

Schulze, I.T. (1972) The structure of influenza virus. II. A model based on the morphology and composition of subviral particles. *Virology* **47**, 181-196

Sieczkarski, S.B. and Whittaker, G.R. (2002) Influenza virus can enter and infect cells in the absence of clathrin-mediated endocytosis. *J. Virol.* **76**, 10455-10464

Skehel, J.J. and Schild, G.C. (1971) The polypeptide composition of influenza A viruses. *Virology* **44**, 396-408

Skehel, J.J., Bayley, P.M., Brown, E.B., Martin, S.R., Waterfield, M.D., White, J.M., Wilson, I.A. and Wiley, D.C. (1982) *Proc. Natl. Acad. Sci. USA* **79**,968-972

Skehel, J.J. and Wiley, D.C. (2000) Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu. Rev. Biochem.* **69**, 531-569

Skehel, J. J., Cross, K., Steinhauer, D. and Wiley, D.C. (2001) Influenza fusion peptides. *Biochem. Soc. Trans.* **29**, 623-626

Skibbens, J.E., Roth, M.G. and Matlin, K.S. (1989) Differential extractability of influenza virus hemagglutinin during intracellular transport in polarized epithelial cells and nonpolar fibroblasts. *J. Cell Biol.* **108**, 821-832

Slemons, R.D. and Easterday, B.C. (1977) Type-A influenza viruses in the feces of migratory waterfowl. *J. Am. Vet. Med. Assoc.* **171**, 947-948

Smith, W., Andrewes, C.H. and Laidlaw, P.P. (1933) A virus obtained from influenza patients. *Lancet*, **ii**, 66-68

Smith, W. and Stuart-Harris, C.H. (1936) Influenza infection of man from the ferret. *Lancet*, **ii**, 121-123

Sowadski, J.M. (1994) Crystallization of membrane proteins. *Curr. Opin. Struct. Biol.* **4**, 761-764

Spaltenstine, A. and Whitesides, G.M. (1991) Polyacrlyamides bearing pendant α-sialoside groups strongly inhibit agglutination of erythrocytes by influenza virus. *J. Am. Chem. Soc.* **113**, 686-687

Stegmann, T., Morselt, H.W.M., Booy, F.P., van Breemen, J.F.L., Scherphof, G. and Wilschut, J. (1987) Functional reconstitution of influenza virus envelopes. *EMBO J.* **6**, 2651-2659

Steinhauer, D.A. (1999) Role of hemagglutinin cleavage for the pathogenicity of influenza virus. *Virology*, **258**, 1-20

Stevens, J., Corper, A.L., Basler, C.F., Taubenberger, J.K., Palese, P. and Wilson, I.A. (2004) Structure of the uncleaved human H1 hemagglutinin from the extinct 1918 influenza virus. *Science* **303**, 1866-1870

Stevens, J., Blixt, O., Tumpey, T.M., Taubenberger, J.K., Paulson, J.C. and Wilson, I. (2006a) Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. *Science* **312**, 404-410

Stevens, J., Blixt, O., Paulson, J.C., Wilson, I.A. (2006b) Glycan microarray technologies: tools to survey host specificity of influenza viruses. *Nat. Rev. Microbiol.* **4**, 857-864

Stieneke-Grober, A., Vey, M., Angliker, H., Shaw, E., Thomas, G., Roberts, C., Klenk, H.D. and Garten, W. (1992) Influenza virus hemagglutinin with multibasic cleavage site is activated by furin, a subtilisin-like endoprotease. *EMBO J.* **11**, 2407-2414

Subbarao, K., Klimov, A., Katz, J., Regnery, H., Lim, W., Hall, H., Perdue, M., Swayne, D., Bender, C., Huang, J., Hemphill, M., Rowe, T., Shaw, M., Xu, X., Fukuda, K. and Cox, N. (1998) Characterization of an avian influenza A (H5N1) virus isolated from a child with a fatal respiratory illness. *Science* **279**, 393-396

Sugrue, R.J., Belshe, R.B. and Hay, A.J. (1990) Palmitoylation of the influenza A virus M2 protein. *Virology* **179**, 51-56

Sugrue, R.J. and Hay, A.J. (1991) Structural characteristics of the M2 protein of influenza A viruses: evidence that it forms a tetrameric channel. *Virology* **180**, 617-624

Suzuki, Y., Matsunaga, M. and Matsumoto, M. (1985) N-Acetylneuraminyl-lactosylceramide, GM3-NeuAc, a new influenza A virus receptor which mediates the adsorption-fusion process of viral infection. Binding specificity of influenza virus A/Aichi/2/68 (H3N2) to membrane-associated GM3 with different molecular species of sialic acid. *J. Biol. Chem.* **260**, 1362-1365

Suzuki, Y., Nagao, Y., Kato, H., Matsumoto, M., Nerome, K., Nakajima, K. and Nobusawa, E. (1986) Human influenza A virus hemagglutinin distinguishes sialyloligosaccharides in membrane-associated gangliosides as its receptor which mediates the adsorption and fusion processes of virus infection. Specificity for oligosaccharides and sialic acids and the sequence to which sialic acid is attached. *J. Biol. Chem.* **261**, 17057-17061

Suzuki, Y., Nakao, T., Ito, T., Watanabe, N., Toda, Y., Xu, G., Suzuki, T., Kobayashi, T., Kimura, Y. and Yamada, A. (1992) Structural determination of gangliosides that bind to influenza A, B, and C viruses by an improved binding assay: strain-specific receptor epitopes in sialo-sugar chains. *Virology*, **189**, 121-131

Suzuki Y. (1994) Gangliosides as influenza virus receptors. Variation of influenza viruses and their recognition of the receptor sialo-sugar chains. *Prog. Lipid Res.* **33**, 429-457

Suzuki, T., Horiike, G., Yamazaki, Y., Kawabe, K., Masuda, H., Miyamoto, D., Matsuda, M., Nishimura, S.I., Yamagata, T., Ito, T., Kida, H., Kawaoka, Y. and Suzuki, Y. (1997) Swine influenza virus strains recognize sialylsugar chains containing the molecular species of sialic acid predominantly present in the swine tracheal epithelium. *FEBS Lett.* **404**, 192-196

Suzuki, Y. (2005) Sialobiology of influenza: molecular mechanism of host range variation of influenza viruses. *Biol. Pharm. Bull.* **28**, 399-408

Swalley, S.E., Baker, B.M., Calder, L.J., Harrison, S.C., Skehel, J.J. and Wiley, D.C. (2004) Full-length influenza hemagglutinin HA_2 refolds into the trimeric low-pH-induced conformation. *Biochemistry* **43**, 5902-5911

Takemoto, D.K., Skehel, J.J. and Wiley, D.C. (1996) A surface plasmon resonance assay for the binding of influenza virus hemagglutinin to its sialic receptor. *Virology* **217**, 452-458

Tamm, L. K. (2003) Hypothesis: spring-loaded boomerang mechanism of influenza hemagglutinin-mediated membrane fusion. *Biochim. Biophys. Acta* **1614**, 14-23

Tatulian, S.A. and Tamm, L.K. (2000) Secondary structure, orientation, oligomerization, and lipid interactions of the transmembrane domain of influenza hemagglutinin. *Biochemistry* **39**, 496-507

Taubenberger, J.K., Reid, A.H., Krafft, A.E., Bijwaard, K.E. and Fanning, T.G. (1997) Initial genetic characterization of the 1918 "Spanish" inflenza virus. *Science* **275**, 1793-1796

Taylor, A.R., Sharp, D.G., Beard, D., Beard, J.W., Dingle, J.H. and Feller, A.E. (1943) Isolation and characterisation of influenza A virus (PR8 strain). *J. Immunolgy.* **47**, 261-282

Taylor, G. (2003) The phase problem. *Acta Cryst.* **D59**, 1881-1890

Unge, T. (1999) Crystallization Methods. In: Protein Crystallisation: Techniques, Strategies and Tips. A Laboratory Manual. T.M. Bergfors, ed. IUL Press, La Jolla, California, *pp.* 7-17

Vaccaro, L., Cross. K.J., Kleinjung, J., Strauss, S.K., Thomas, D.J., Wharton, S.A., Skehel, J.J. and Fraternali, F. (2005) Plasticity of influenza haemagglutinin fusion peptides and their interaction with lipid bilayers. *Biophys. J.* **88**, 25-36

Varki, A. (1992) Diversity in the sialic acids. *Glycobiology* **2**, 25-40

Varki, A. (1997) Sialic acids as ligands in recognition phenomena. *FASEB J.* **11**, 248-255

Veit, M., Kretzschmar, E., Kuroda, K., Garten, W., Schmidt, M.F., Klenk, H.D. and Rott, R. (1991) Site-specific mutagenesis identifies three cysteine residues in the cytoplasmic tail as acylation sites of influenza virus hemagglutinin. *J. Virol.* **65**, 2491-2500

Verhoeyen, M., Fanq, R., Jou, W. M., Devos, R., Huylebroeck, D., Saman, E. and Fiers, W. (1980) Antigenic drift between the haemagglutinin of the Hong Kong influenza strains A/Aichi/2/68 and Victoria/3/75. *Nature* **286**, 771-776

Vines, A., Wells, K., Matrosovich, M., Castrucci, M.R., Ito, T. and Kawaoka, Y. (1998) The role of influenza A virus hemagglutinin residues 226 and 228 in receptor specificity and host range restriction. *J. Virol.* **72**, 7626-7631

Wallace A C, Laskowski R A & Thornton J M (1995). LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. *Prot. Eng.* **8**, 127-134.

Warren, L (1959) The thiobarbituric acid assay of sialic acids. *J. Biol. Chem.* **234**, 1971-1975

Watowich, S.J., Skehel, J.J. and Wiley, D.C. (1994) Crystal structures of influenza virus hamagglutinin in complex with high-affinity receptor analogs. *Structure* **2**, 719-731

Webster, R.G., Yakhno, M., Hinshaw, V.S., Bean, W.J. and Murti, K.G. (1978) Intestinal influenza: replication and characterization of influenza viruses in ducks. *Virology* **84**, 268-278

Webster, R.G., Bean, W.J., Gorman, O.T., Chambers, T.M. and Kawaoka, Y. (1992) Evolution and ecology of influenza A viruses. *Microbiol. Rev.* **56**, 152-179

Webster, R.G. (1997) Influenza virus, transmission between species and relevance to emergence of the next pandemic. *Arch. Virol. Suppl.* **13**, 105-113

Weis, W., Brown, J.H., Cusack, S., Paulson, J.C., Skehel, J.J. and Wiley, D.C. (1988) Structure of the influenza virus haemagglutinin complexed with its receptor, sialic acid. *Nature* **333**, 426-431

Weis, W.I., Brunger, A.T., Skehel, J.J. and Wiley, D.C. (1990) Refinement of the influenza virus hemagglutinin by simulated annealing. *J. Mol. Biol.* **212**, 737-761

Wharton, S.A., Skehel, J.J. and Wiley, D.C. (1986) Studies of influenza haemagglutinin-mediated membrane fusion. *Virology* **149**, 27-35

White, J., Matlin, K. and Helenius, A. (1981) Cell fusion by Semliki Forest, influenza, and vesicular stomatitis viruses. *J. Cell Biol.* **89**, 674-679

WHO (1953) EXPERT Committee on Influenza, first report. *World Health Organ. Tech. Rep. Ser.* **6**, 1-32

Wiley, D.C. and Skehel, J.J. (1987) The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. *Annu. Rev. Biochem.* **56**, 365-394

Wilson, I.A., Skehel, J.J. and Wiley, D.C. (1981) Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. *Nature* **289**,366-73

Wooh, J.W., Kidd, R.D., Martin, J.L. and Kobe, B. (2003) *Acta Cryst.* **D59**, 769-772

Wu, W. and Air, G.M. (2004) Binding of influenza viruses to sialic acids: reassortant viruses with A/NWS/33 hemagglutinin bind to alpha2,8-linked sialic acid. *Virology* **325**, 340-350

Xu, G., Suzuki, T., Tahara, H., Kiso, M., Hasegawa, A. and Suzuki, Y. (1994) Specificity of sialyl-sugar chain mediated recognition by the hemagglutinin of human influenza B virus isolates. *J. Biochem.* **115**, 202-207

Yamada, S., Suzuki, Y., Suzuki, T., Le, M.Q., Nidom, C.A., Sakai-Tagawa, Y., Muramoto, Y., Ito, M., Kiso, M., Horimoto, T., Shinya, K., Sawada, T., Kiso, M., Usui, T., Murata, T., Lin, Y., Hay, A., Haire, L.F., Stevens, D.J., Russell, R.J., Gamblin, S.J., Skehel, J.J. and Kawaoka, Y. (2006) Haemagglutinin mutations responsible for the binding of H5N1 influenza A viruses to human-type receptors. *Nature* **444**, 378-382

Zebedee, S.L. and Lamb, R.A. (1988) Influenza A virus M2 protein: monoclonal antibody restriction of virus growth and detection of M2 in virions. *J. Virol.* **62**, 2762-2772

Zhang, J., Pekosz, A. and Lamb, R.A. (2000) Influenza virus assembly and lipid raft microdomains: a role for the cytoplasmic tails of the spike glycoproteins. *J. Virol.* **74**, 4634-4644

Appendix I

Comparison of the crystal structures of HAs of different subtypes and hosts of origin

	R.m.s.d. (Å)							
	H3 avian HA	H3 human HA	H5 avian HA	H7 avian HA	H9 swine HA	H1 1918 HA	H1 human HA	H1 swine HA
H3 avian HA		0.50	2.02 (410)	1.47	1.98 (433)	2.08 (397)	2.11 (408)	2.02 (396)
H3 human HA			2.08 (401)	1.46	2.14 (430)	2.15 (408)	2.11 (399)	2.08 (399)
H5 avian HA				2.14 (340)	1.38	1.03	1.08	0.90
H7 avian HA					2.34 (407)	2.20 (353)	2.13 (337)	2.24 (354)
H9 swine HA						1.46	1.59	1.54
H1 1918 HA							0.64	0.56
H1 human HA								0.57
H1 swine HA								

Table 7. Structural comparison of the HAs from different subtypes and hosts of origin

Overall *r.m.s.d.* values for the pairwise comparison of the C_α positions of monomers of the HAs from different subtypes and hosts of origin. The different virus strains are: Duck/Ukraine/1/63 (H3 avian), A/Aichi/2/68 (H3 human), Duck/Singapore/3/97 (H5 avian), Turkey/Italy/02 (H7 avian), Swine/Hong Kong/9/98 (H9 swine), Puerto Rico/8/34 (H1 human) and Swine/Iowa/15/30 (H1 swine). See text for details about the different HAs, namely *1.5.2*, *1.6.6*, *1.6.7* and *1.6.9*. The gene for the 1918 HA was made synthetically (Gamblin et al., 2004).

The more recently determined crystal structures of avian HAs from H5N1 viruses, the A/Vietnam/1203/04 HA and A/Vietnam/1194/04 HA, are very similar to the indicated H5 avian HA, A/duck/Singapore/3/97, with a *r.m.s.d.* of 0.46 Å and 1.70 Å, respectively (Stevens et al., 2006a; Yamada et al., 2006). Numbers in parentheses refer to the number of amino acids that could be included in the comparisons.

[Adapted from: Russell et al., 2004]

Appendix II

NSC definition for the L226Q HA in the complexes with LSTa and LSTc

	L226Q HA - LSTa						L226Q HA - LSTc					
	Matrix1			Matrix 2			Matrix1			Matrix 2		
Rotation	0.424434	-0.553326	-0.716719	0.424410	-0.846569	0.321243	0.424705	-0.554606	-0.715568	0.424325	-0.846018	0.322804
	-0.849656	0.030199	-0.526472	-0.547494	0.042651	0.835722	-0.849200	0.029938	-0.527222	-0.547908	0.043941	0.835384
	0.312954	0.832418	-0.457319	-0.721198	-0.530568	-0.445390	0.313824	0.831574	-0.458257	-0.720934	-0.531341	-0.444895
Translation	51.662	83.749	-54.331	66.326	70.557	56.884	51.569	83.391	-54.116	66.031	70.178	56.702

Table 8. Transformation matrices for the generation of the L226Q HA monomers 2 and 3 from monomer 1 by application of the three-fold NCS operators

The NCS definition of the L226Q HA trimer, in complex with LSTa and LSTc, is given by transformation matrices. Matrix 1 transforms monomer 1 into monomer 2, and Matrix 2 monomer 1 into monomer 3, in the context of a L226Q HA trimer having a NCS axis. Monomer 1 corresponds to the L226Q HA monomer containing the RBS1 (to which the LSTa *trans* conformer binds), monomer 2 harbours the ligand-free L226Q HA RBS3, and monomer 3 corresponds to the L226Q HA monomer containing the RBS2 (to which the LSTa *cis* conformer binds) (see *4.6*). These transformation matrices were determined using the program LSQMAN (USF suite) (see *2.2.4.4.1*). The transformation operators were derived from the alignment of the Cα-chain atoms (503 atoms) of each monomer.

Appendix III

Interatomic interactions between crystallographic symmetry related L226Q HA trimers

Table 9: Lattice contact interactions in the crystal of L226Q HA in complex with LSTa

A list of the interatomic distances between residues of crystallographic symmetry related L226Q HA trimers forming hydrogen bonds is shown.

Contact region A		
L226Q HA trimer 1 at (*x*, *y*, *z*)	**Distance (Å)**	**L226Q HA trimer 2 at (*-y*, *x*, *z+1/4*)**
HA_1 monomer 3		***HA_2 monomer 3***
Trp 127 (m/c CO)	3.19	Arg 25 (NH2)
Thr 128 (OG1)	3.94	His 26 (m/c NH)
Thr 128 (OG1)	3.43	His 26 (m/c CO)
Gln 132 (OE1)	3.94	Arg 25 (m/c NH1)
Ser 157 (OG)	3.91	Asn 135 (ND2)
Ser 157 (m/c CO)	2.65	Asn 135 (ND2)
Gly 158 (m/c NH)	3.39	Glu 11 (OE1)
Thr 160 (OG1)	3.07	Asn 135 (ND2)
Val 163 (m/c CO)	2.86	Gln 27 (NE2)
Asn 165 (m/c NH)	3.32	Gln 27 (OE1)
Asn 165 (m/c NH)	3.55	Gln 27 (NE2)
Asn 165 (m/c CO)	3.75	Gln 27 (OE1)
GlcNAc 339 (O6) [at Asn 165]	3.92	Gly 31 (m/c NH)
GlcNAc 340 (O5) [at Asn 165]	3.95	SER 29 (m/c CO)

Table 9 (cont.)

Contact region A		
L226Q HA trimer 1 at (*x*, *y*, *z*)	**Distance (Å)**	**L226Q HA trimer 1 at (*-y*, *x*, *z+1/4*)**
HA$_1$ monomer 3		***HA$_1$ monomer 3***
Gln 132 (m/c NH)	3.65	Glu 325 (OE2)
Gln 132 (OE1)	3.89	Glu 325 (OE1)
Gln 132 (m/c CO)	3.77	Glu 325 (m/c CO)
SER 145 (OG)	3.78	Thr 328* (m/c CO)
HA$_1$ monomer 2		***HA$_1$ monomer 3***
Gln 189 (OE1)	3.88	Asn 8* (OD1)
Gln 189 (NE2)	3.33	Asn 8* (ND2)
HA$_1$ monomer 2		***HA$_2$ monomer 3***
Ser 193 (OG)	3.95	Lys 174* (NZ)
Contact region B		
L226Q HA trimer 1 at (*x*, *y*, *z*)	**Distance (Å)**	**L226Q HA trimer 2 at (*y*, *-x*, *z+3/4*)**
HA$_1$ monomer 1		***HA$_2$ monomer 2***
Asn 53 (ND2)	3.19	Asn 146 (ND2)
Arg 57 (NH1)	2.24	Ser 29 (m/c CO)
Arg 57 (NH1)	3.89	Glu 30 (m/c NH)

Arg 57 (NH1)	3.58	Glu 30 (m/c CO)
Arg 57 (NH1)	3.61	Gly 31 (m/c NH)
Asp 60 (m/c NH)	3.84	Thr 32 (m/c CO)
Asp 60 (m/c NH)	3.42	Thr 32 (OG1)
Asp 60 (m/c CO)	2.70	Thr 32 (OG1)
Asp 63 (OD2)	3.29	Arg 25 (NH1)
Asp 63 (OD2)	3.08	Arg 25 (NH2)
Asp 63 (OD2)	3.29	Arg 25 (NH1) [s]
Asp 63 (OD2)	3.08	Arg 25 (NH2) [s]
Val 78 (m/c CO)	3.57	Gln 27 (OE1)
Val 78 (m/c CO)	3.52	Gln 27 (NE2)
Glu 82 (OE2)	3.09	Gln 27 (NE2)
Lys 92 (NZ)	2.99	Gly 16 (m/c CO)
HA_1 monomer 1		***HA_1 monomer 2***
Asn 96 (OD1)	3.89	Glu 325 (m/c CO)

The interaction distances were determined using the program CONTACT (CCP4 suite) and checked manually in O (Jones et al., 1991). These interactions were also identified by the PISA server (Krissinel and Henrick, 2007), which has additionally assigned salt bridges. Contact regions A and B are described in *4.6.3* and shown graphically in Figure 34. Monomers 1 and 3 correspond, respectively, to the L226Q HA monomers harbouring the RBS1, which binds the LSTa *trans* conformer, and RBS2, which binds the LSTa *cis* conformer. Monomer 2 corresponds to the L226Q HA monomer containing the RBS3, which does not bind either LSTa or LSTc (see *4.6.3*). The atoms in the left column correspond to the reference L226Q HA trimer and those in the right column correspond to the symmetry related trimer. In the latter column the symmetry operator is also shown.

The superscript *s* indicates a potential salt bridge.

* These residues show high *B* values (> 65 $Å^2$) and therefore the assigned interactions have a high level of error.

Appendix IV

Comparison of the interatomic interactions of LSTa and LSTc with L226Q HA and HAs of different subtypes and hosts of origin

Table 10. Interatomic interaction distances between the receptor analogue LSTa and the HAs of different subtypes and hosts of origin

The interatomic distances listed correspond to potential hydrogen bonds of the Sia and Gal-2 residues of LSTa with L226Q HA, X-31 HA, H3 avian HA, H5 avian HA, H7 avian HA, H9 swine HA, H1 human HA and H1 swine HA (see: Figures 7-13; *1.6.7.1*). H3 numbering was applied in all cases.

				Bond distance (Å)								
LSTa		**RBS**		**L226Q HA**		**X-31 HA**	**H3 avian HA**	**H5 avian HA**	**H7 avian HA**	**H9 swine HA**	**H1 human HA**	**H1 swine HA**
Residue	**Atom**	**Atom**	**Residue**	***Trans*** $_{RBS1}$	***Cis*** $_{RBS2}$	***Cis***	***Trans***	***Trans***	***Trans***	***Cis***	***Trans***	**-**
Sia	O1A	m/c NH	137	2.7	2.8	2.9	2.7	2.7	3.2	2.9	2.7	2.6
	O1B	m/c NH	137	-	-	-	-	-	-	-	-	-
	O1A	OG	S137	x	x	x	-	-	3.2	-	x	x
	O1A	OG	S136	-	3.7	3.5	3.1	3.4	3.5	-	x	x
	O1B	OG	S136	3.0	2.7	2.8	2.6	2.5	2.7	3.3	x	x
	O1A	OG1	T136	x	x	x	x	x	x	x	-	-
	O1B	OG1	T136	x	x	x	x	x	x	x	2.5	2.7

Table 10 (cont.)

LSTa		RBS		Bond distance (Å)								
				L226Q HA		X-31 HA	H3 avian HA	H5 avian HA	H7 avian HA	H9 swine HA	H1 human HA	H1 swine HA
Residue	**Atom**	**Atom**	**Residue**	***Trans*** $_{RBS1}$	***Cis*** $_{RBS2}$	***Cis***	***Trans***	***Trans***	***Trans***	***Cis***	***Trans***	-
Sia	N5	m/c CO	135	2.9	3.0	3.0	2.8	3.1	3.3	2.9	3.0	2.9
	O8	OE1	Q226	2.7	3.2	x	2.9	3.3	2.8	x	3.0	3.1
	O8	OH	Y98	3.3	2.8	2.9	3.1	3.0	2.7	2.9	2.7	3.2
	O8	Wat1						2.9				
	O9	OH	Y98	2.6	3.2	2.7	3.1	2.8	2.7	3.3	3.0	-
	O9	OE1	E190	2.2	2.2	3.0	2.7	2.8	2.8	x	-	x
	O9	OE2	E190	-	3.6	2.9	3.5*	-	-	x	2.5	x
	O9	NE2	H183	3.8	3.9	3.0	3.1	2.8	3.3	x	3.1	3.4
	O9	OG	S228	3.2	3.2	2.8	x	x	x	-	x	x
	O9	Wat1						3.1		2.9		

Table 10 (cont.)

LSTa		RBS		Bond distance (Å)								
				L226Q HA		X-31 HA	H3 avian HA	H5 avian HA	H7 avian HA	H9 swine HA	H1 human HA	H1 swine HA
Residue	Atom	Atom	Residue	*Trans* RBS1	*Cis* RBS2	*Cis*	*Trans*	*Trans*	*Trans*	*Cis*	*Trans*	-
Sia	Wat1	m/c NH	228					2.8		2.7		
	O1A	NE2	Q226	-	-	x	3.2	3.4	3.4	x	-	-
	O1B	NE2	Q226	3.4	3.5	x	3.0	3.6	3.3	x	3.5	3.7
	O1A	ND2	N137	2.8	3.2	3.0	x	x	x	x	x	x
Gal-2	O3	NE2	Q226	(5.5)	(5.3)	x	3.5	3.7	3.5	x	3.9	-
	O4	OE1	Q226	3.7	-	x	3.5	3.0	3.0	x	3.3	-
	O4	NE2	Q226	-	3.6	x	3.3	-	-	x	-	-
	O4	ND2	N137	-	3.0	-	x	x	x	x	x	x
	O4	Wat1									2.8	

Table 10 (cont.)

LSTa		RBS		Bond distance (Å)								
				L226Q HA		X-31 HA	H3 avian HA	H5 avian HA	H7 avian HA	H9 swine HA	H1 human HA	H1 swine HA
Residue	Atom	Atom	Residue	*Trans* RBS1	*Cis* RBS2	*Cis*	*Trans*	*Trans*	*Trans*	*Cis*	*Trans*	-
Gal-2	O4	Wat2									2.9	
	Wat1	NZ	K222	x	x	x	x	-	x	x	3.3	-
	Wat1	m/c CO	225								2.9	
	Wat2	m/c NH	227	-	-	-	x	-	-	-	3.4	
	O6	m/c CO	225	-	3.0	2.8	-	-	-	2.6	-	-

The average distances are given except in the cases where LSTa binds in only one of the NCS-related RBSs of the HA trimer. Distances were assigned using the molecular graphics program O (Jones et al., 1991). The coordinates for X-31 HA-LSTa and H7 avian HA-LSTa were provided by M. B. Eisen and R. J. Russell, respectively. The coordinates for LSTa in complex with H3 avian HA, H5 avian HA, H9 swine HA, H1 human HA and H1 swine HA were obtained from RCSB/PDB, accession codes 1mqm, 1jsn, 1jsh, 1rvx and 1rv0, respectively. - : the interaction is not observed in the designated HA. x : the position of the designated HA does not contain the specified amino acid residue. * : only assigned for two monomers. The configuration of the α2,3 Sia-Gal glycosidic linkage is also indicated (See Figures 7-13).

Table 11. Interatomic interaction distances between the receptor analogue LSTc and the HAs of different subtypes and hosts of origin

The interatomic distances listed correspond to potential hydrogen bonds of the Sia of LSTc with L226Q HA, X-31 HA, H3 avian HA, H5 avian HA, H9 swine HA, H1 human HA and H1 swine HA (see Figures 7-13, *1.6.7.1*). H3 numbering was applied in all cases.

				Bond distance (Å)						
LSTc		**RBS**		**L226Q HA**	**X-31 HA**	**H3 avian HA**	**H5 avian HA**	**H9 swine HA**	**H1 human HA**	**H1 swine HA**
Residue	**Atom**	**Atom**	**Residue**	-	*Cis*	*Cis*	-	*Cis*	*Cis*	*Cis*
Sia	O1A	m/c NH	137	2.8	2.9	3.0	2.9	2.8	3.0	2.8
	O1B	m/c NH	137	-	-	-	-	-	3.3	3.2
	O1A	OG	S137	x	x	-	-	-	x	x
	O1A	OG	S136	3.5	3.1	3.5	2.6	-	x	x
	O1B	OG	S136	2.7	2.8	3.0	2.9	3.1	x	x
	O1A	OG1	T136	x	x	x	x	x	-	-
	O1B	OG1	T136	x	x	x	x	x	2.6	2.6

Table 11 (cont.)

LSTc		RBS		Bond distance (Å)						
				L226Q HA	**X-31 HA**	**H3 avian HA**	**H5 avian HA**	**H9 swine HA**	**H1 human HA**	**H1 swine HA**
Residue	**Atom**	**Atom**	**Residue**	-	*Cis*	*Cis*	-	*Cis*	*Cis*	*Cis*
Sia	N5	m/c CO	135	2.8	3.0	3.3	2.9	2.8	3.0	3.4
	O8	OE1	Q226	2.8	x	3.2	3.3	x	3.1	2.8
	O8	OH	Y98	3.1	2.7	3.2	2.9	2.8	2.6	2.7
	O9	OH	Y98	2.7	2.8	2.9	2.7	3.5	3.0	3.1
	O9	OE1	E190	2.1	2.8*	3.1	2.7	x	3.2	x
	O9	OE2	E190	-	3.0	-	3.2	x	2.6	x
	O9	NE2	H183	3.2	3.1	3.5	3.0	x	3.3	3.3

Table 11 (cont.)

LSTc		RBS		Bond distance (Å)						
				L226Q HA	**X-31 HA**	**H3 avian HA**	**H5 avian HA**	**H9 swine HA**	**H1 human HA**	**H1 swine HA**
Residue	**Atom**	**Atom**	**Residue**	-	*Cis*	*Cis*	-	*Cis*	*Cis*	*Cis*
Sia	O9	OG	S228	3.1	2.9	x	x	-	x	x
	O9	Wat1						2.9		
	Wat1	m/c NH	228					3.2		
	O1A	NE2	Q226	-	x	-	-	x	-	-
	O1B	NE2	Q226	3.2	x	2.9	3.3	x	3.7	3.5
	O1A	ND2	N137	3.3	-	x	x	x	x	x

The average distances are given except in the cases where LSTc binds in only one of the NCS-related RBSs of the HA trimer. Distances were assigned using the molecular graphics program O (Jones et al., 1991). The coordinates for X-31 HA-LSTc were provided by M. B. Eisen. The coordinates for LSTc in complex with H3 avian HA, H5 avian HA, H9 swine HA, H1 human HA and H1 swine HA were obtained from RCSB/PDB, accession codes 1mqn, 1jso, 1jsi, 1rvz and 1rvt, respectively. - : the interaction is not observed. x : the position does not contain the specified amino acid residue. * : only assigned for two monomers. The configuration of the α2,6 Sia-Gal glycosidic linkage is also indicated (see Figures 7-13).

Appendix V

Standard three and one letter amino acid codes are used throughout the text. The table below describes the two alternative codes for the designation of amino acids.

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

Standard three and one letter amino acid codes