X-ray Crystallographic Studies of Viral and Human Membrane Glycoproteins

A. Carfi,[1,2] R. Gaudet,[1,3] Y. Ha,[1,3] J. Hennecke,[1,3] K.-H. Lee,[1,2] D. C. Wiley[1,2,3]

[1] *Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA, U.S.A.*

[2] *Department of Medicine, The Children's Hospital, Boston, MA, U.S.A.*

[3] *Howard Hughes Medical Institute, Cambridge, MA, U.S.A.*

At the APS, we collected data sets on five projects in 2000: human herpes simplex glycoprotein D bound to the human receptor HveA, the complex between human cytomegalovirus protein US2 and MHC class I molecule HLA-A2, a number of influenza haemagglutinin crystal structures from avian and swine viruses, the three-dimensional structure of a ternary complex between human T-cell receptor HA1.7, influenza virus haemagglutinin peptide HA 306-318 and human MHC class II protein HLA-DR4; and the structure of a Human Insulin Peptide/HLA-DQ8 complex related to susceptibility to Type 1 Diabetes.

1. Human herpes simplex glycoprotein D bound to the human receptor HveA

An essential step in herpes simplex virus (HSV) infection is mediated by binding of the viral envelope glycoprotein D (gD) to cell surface receptors. We have determined the x-ray structures of a soluble, truncated ectodomain of gD in complex with the ectodomain of HveA, an HSV cellular receptor expressed on the surface of human T lymphocytes and a member of the tumor necrosis factor receptor superfamily (TNFR).

A native data set and a $K_2Pt(CN)_4$ derivative data set were collected at BioCars beamline 14-BM-C and a data set at the selenium absorption-peak wavelength was collected from a crystal containing selenomethionine-substituted gD at the 14-BM-D beam line (Table 1). The platinum derivative and the anomalous signal from the selenomethionines provided together enough phase information to solve the structure at 2.9 Å resolution. Subsequently, the structure was refined at 2.65 Å resolution using a data set collected at the CHESS A1 beamline on a crystal with smaller cell parameters.

Unexpectedly, the structure reveals a V-like immunoglobulin (Ig) fold at the core of gD that is closely related to cellular adhesion molecules and that is flanked by large N- and C-terminal extensions. The HveA binding site is entirely on a 32-residue N-terminal segment that forms a hairpin at one extreme edge of the gD molecule, rather than being a typical surface or a pocket that is assembled from many parts of the gD sequence.

Two bound anions suggest possible binding sites for another gD receptor, a 3-O-sulfonated heparin sulfate. The atomic details of the gD-HveA interface and its proximity to one of the ion binding sites and to the putative binding site for another receptor HveC suggest how these interactions might all be inhibited by a single small molecule.

2. Complex between human cytomegalovirus protein US2 and MHC class I molecule HLA-A2

A common feature of many persistent viruses is the ability to subvert MHC class I antigen presentation. The human cytomegalovirus (HCMV) US2 glycoprotein targets newly synthesized MHC class I heavy chains for rapid dislocation from the endoplasmic reticulum (ER) to the cytosol, where they are degraded by the proteasome. The purpose of this project was to gain insight into US2 structure and function.

A native data set was collected of the US2/HLA-A2/Tax peptide crystals at BIOCARS beamline 14-BM-D. The data from one of them was used to determine the structure of the complex by molecular replacement using free HLA-A2/Tax as a search model. The structure has been refined and analyzed. US2 associates with HLA-A2 at the junction of the peptide-binding region and the alpha3 domain. Examining ER dislocation of class I molecules biochemically with mutant heavy chains confirms the importance of this binding site *in vivo*. Unexpectedly, the US2 ER-lumenal domain has an immunoglobulin-like (Ig) fold. A US2 structure-

Table 1.

	Native	**$K_2Pt(CN)_4$**	**SeMet**
	Data Collection		
λ (Å)	1.0	1.0	0.97910
Space group	P3(1)21	P3(1)21	P3(1)21
Unit cell / a b,c (Å)	130.91, 130.91, 82.56	131.09, 131.09, 82.56	131.48, 131.48, 82.67
Resolution (Å)	2.90	3.00	3.80
Mosaicity	0.16	0.19	
Beamline	14-BM-C/BIOCARS	14-BM-C/BIOCARS	14-BM-C/BIOCARS
Total reflections	264,902	301,942	80,740
Unique reflections	18,404	16,621	8.260
Completeness (%)	99.9 (100)	99.6 (99.9)	99.7 (99.8)
R_{sym} (%)[1]	7.0 (41.3)	6.8 (42.7)	5.6 (11.4)
R_{deriv} (%)[2]		0.09	0.14
Phasing Power (acentric)[3]		1.5	0.6
Figure of Merit	0.47		

based sequence alignment reveals that seven HCMV proteins, at least three of which interfere with cell surface expression of MHC class I molecules, share the same fold as US2. The structure provides the first example of a complex between class I and a protein that binds to it in *cis*. This structure also has implications for the mechanism by which US2 recognizes and targets class I for proteasome degradation.

Resolution	40.0-2.20 Å (2.28-2.20 Å)
Number of reflections	139,486 (13,849)
Average I/σI	14.2 (3.3)
Completeness (%)	98.0 (97.1)
R_{sym} (%)	4.5 (28.7)
R_{crys} (%)	20.5 (30.0)[4]
R_{free} (%)	24.0 (34.8)[4]

[4] R_{cryst} and $R_{free} = \Sigma h||F(h)obs|-|F(h)calc||/\Sigma h|F(h)obs|$ for reflections in the working and test sets, respectively. Numbers in parentheses are for final shell.

3. Influenza haemagglutinin crystal structures

We have collected native and heavy atom derivative data of hemagglutinins (HA) of different influenza A virus subtypes (H1, H2, H3, H5 and H9) isolated from human, swine and avian hosts at APS BIOCARS 14-BM-B and -C beamlines. We subsequently determined their atomic structures. We observed a major conformational difference at the membrane distal domain that rotates by about 20 degrees. Correlating the difference with sequence motifs allows us to predict that there are two HA structural classes and to hypothesize a structural basis for the origin of influenza virus HA subtypes. The new structures also suggest specific residues in low-pH triggering of membrane fusion activity of H5 and H9 HAs.

Emerging influenza pandemics have been accompanied by the evolution of receptor binding specificity from the preference of avian viruses for sialic acid receptors in α2,3 linkage to the preference of human viruses for α2,6 linkages. We collected data sets of HAs complexed with human- and avian-like receptor analogs at BIOCARS. The complex crystal structures showed how avian and human receptors are distinguished by atomic contacts at the linkage and by the "openness" of the HA binding site, suggesting how the receptor binding sites of HAs from avian viruses evolve the potential to emerge in humans.

4. Three-dimensional structure of a complex between human T-cell receptor HA1.7, influenza virus haemagglutinin peptide HA 306-318 and human MHC class II protein HLA-DR4

We recently solved the structure of the human T-cell receptor HA1.7 in complex with the human MHC class II protein HLA-DR1 presenting the influenza virus haemagglutinin peptide HA 306-318.[1] It has been shown by T-cell activation assays that TCR HA1.7 also recognizes the HA peptide presented by HLA-DR4. Although all residues on the surface of DR1 and DR4 that potentially can be contacted by the TCR are identical—allelic differences in the DR β chain are located deep in the peptide binding groove and in the β2 domain—other TCRs however are not cross reactive and only recognize HA presented by either DR1 or DR4.[2,3] To get a structural insight into this observation, we solved the structure of the TCR HA1.7/HLA-DR4/HA complex.

Crystals were grown from 16% PEG8000, 1 M NaCl, 100 mM HEPES pH 7.0 and frozen using 20% glycerol as a cryoprotectant. The complex crystallized in spacegroup C2 with cell dimensions of with a = 143.8 Å, b = 73.3 Å, c = 123.0 Å, β = 108.5°. Data of this crystal have been measured on the BIOCARS 14-BM-C-station (250 images, 1.0 degree oscillation per image). Processing of the data with DENZO and SCALEPACK gave a native dataset with the following statistical values:

Resolution (Å)	20.0-2.4 (2.49-2.40)
Mosaicity (deg.)	1.1
No. unique reflections	46705
Multiplicity	4.5 (4.2)
Completeness	96.9 (96.7)
Average I/σI	20.6 (4.3)
R_{merge} (%)	5.8 (33.0)

The dataset has been used to determine the structure of the complex by molecular replacement. The structure has been refined to $R_{work} = 0.208$ and $R_{free} = 0.245$.

5. Structure of a human insulin peptide/HLA-DQ8 complex and susceptibility to type 1 diabetes

The class II major histocompatibility complex (MHCII) glycoproteins HLA-DQ8 and HLA-DQ2 in humans and I-A^{g7} in the nonobese diabetic mouse (NOD) are the major risk factors for increased susceptibility to type 1 diabetes. The purpose of the project was to see structural similarities and differences in MHCII/peptide interactions among high-risk and low-risk alleles, such as human DQ2 and murine I-A^{g7}, and DRs, respectively.

X-ray diffraction data for DQ8/insulinB complex was collected from a single crystal to 2.4 Å resolution at the BIOCARS station 14-BM-C using 1 Å wavelength x-rays and Quantum CCD detectors. The space group is $P2_1$ with unit cell dimensions a = 66.05 Å, b = 42.92 Å, c = 87.60 Å, β = 102.5°.

Initial phases were obtained by molecular replacement (MR) using AmoRe in the resolution range of 10 – 3.0 Å with the I-A^k α-chain (PDB accession code 1IAK) and the DR1 β-chain (PDB accession code 1DHL) as search models, separately. The model was refined by iterative rounds of positional and group B-factor refinement and manual model building with data between 30 Å and 2.4 Å applying a bulk solvent correction and overall anisotropic B factor correction. The final model includes 378 amino acids, one carbohydrate residue, and 73 waters (α chain: 2 – 181; β chain: 3 – 104; 112-192, insulinB; 1 linker glycine).

Number of independent reflections	18183
Resolution range	30 - 2.4 Å
Completeness	95%
<I/σ>	8
R_{merge}	10.3% (36.2%)[a]
R_{free}	25.8% (31.1%)[b]
R_{work}	22.4% (31.5%)[b]

[a] Statistics in 2.4-2.49 Å shell.

[b] Statistics in 2.4-2.42 Å shell. $R_{free} = (\Sigma h \mid Fo - Fc \mid / \Sigma h\ Fo)$, $\forall h \in \{\text{free set}\}$; $R_{work} = (\Sigma h \mid Fo - Fc \mid / \Sigma h\ Fo)$, $\forall h \in \{\text{working set}\}$. $R_{merge} = (\Sigma hkl \mid I - \langle I \rangle \mid / (\Sigma hkl \langle I \rangle)$, $\forall hkl \in \{\text{independent Miller indices}\}$.

References

[1] J. Hennecke, A Carfi, D.C. and Wiley, EMBO J. **19**, 5611-5624 (2000).

[2] X.T. Fu, C.P. Bono, S.L. Woulfe, C Swearingen, N.L. Summers, F. Sinigaglia, A. Sette, B.D. Schwartz, and R.W. Karr, J. Exp. Med. **181**, 915-926 (1995).

[3] L.R. Wedderburn, S.J.M. Searle, A.R. Rees, J.R. Lamb, and M.J. Owen, Eur. J. Immunol. **25**, 1654-1662 (1995).