Acta Crystallographica Section D
**Biological Crystallography**
ISSN 0907-4449

# X-ray crystallographic determination of the structure of the influenza C virus haemagglutinin-esterase-fusion glycoprotein

**Xiaodong Zhang,[a,b]† Peter B. Rosenthal,[a]‡ Frank Formanowski,[c]§ Wolfgang Fitz,[d]¶ Chi-Huey Wong,[d] H. Meier-Ewert,[c] John J. Skehel[e] and Don C. Wiley[a,b]***

[a]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA, [b]Howard Hughes Medical Institute, Harvard University, Cambridge, MA 02138, USA, [c]Abteilung für Virologie, Technical University of Munich, Biedersteiner Strasse 29, 80802 Munich 40, Germany, [d]Department of Chemistry, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, CA 92037, USA, and [e]National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, England

† Present address: Molecular Structure and Function Laboratory, Imperial Cancer Research Fund, London WC2A 3PX, England.
‡ Present address: MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England.
§ Present address: Roche des Roches, 1785 Cressier, Switzerland.
¶ Present address: Quest International, Bussum, The Netherlands.

Correspondence e-mail: dcwadmin@crystal.harvard.edu



Received 11 September 1998
Accepted 5 January 1999

**PDB Reference:** haemagglutinin-esterase-fusion protein, 1flc.

The structure of the haemagglutinin-esterase-fusion (HEF) glycoprotein from influenza C virus has been determined to 3.2 Å resolution by X-ray crystallography. A synthetic mercury-containing esterase inhibitor and receptor analogue, 9-acetamidosialic acid $\alpha$-thiomethylmercuryglycoside, was designed as the single isomorphous heavy-atom derivative. The asymmetric unit of one crystal form (form I; $P4_322$, $a = b = 155.4$, $c = 414.4$ Å) contained an HEF trimer. Six mercury sites identifying the three haemagglutination and three esterase sites were located by difference Patterson map analysis of a 6.5 Å resolution derivative data set. These positions defined the molecular threefold-symmetry axis of the HEF trimer. A molecular envelope was defined by averaging a 7.0 Å resolution electron-density map, phased by single isomorphous replacement (SIR), about the non-crystallographic threefold-symmetry axis. Iterative non-crystallographic symmetry averaging in real space, solvent flattening and histogram matching were used to extend the phases to 3.5 Å resolution. Molecular replacement of the model into a second crystal form (form II; $P4_32_12$, $a = b = 217.4$, $c = 421.4$ Å) containing two HEF trimers per asymmetric unit permitted iterative ninefold averaging of the electron density. The 3.5 Å electron-density map allowed an unambiguous tracing of the polypeptide chain and identification of N-linked carbohydrates. The model has been refined by least squares to 3.2 Å resolution ($R_{free}$ = 26.7%).

## 1. Introduction

Influenza C virus is a lipid-enveloped orthomyxovirus which causes respiratory infections in humans. Many strains of influenza C virus appear to circulate simultaneously, causing occasional outbreaks of disease, especially in children (Katagiri *et al.*, 1983; Air & Compans, 1983), but not the familiar periodic epidemics and pandemics caused by influenza A and B viruses (Katagiri *et al.*, 1983; Air & Compans, 1983). The virus membrane contains multiple copies of a single 225 kDa trimeric glycoprotein arranged in an open hexagonal array (Herrler *et al.*, 1981; Hewat *et al.*, 1984). This haemagglutinin-esterase-fusion glycoprotein (HEF) has three activities essential for viral infectivity: receptor binding, receptor-destroying enzyme activity and membrane fusion. Like the HA glycoprotein of influenza A and B viruses, HEF is synthesized as a single-chain precursor $HEF_0$, which is cleaved post-translationally into the disulfide-linked polypeptides $HEF_1$ and $HEF_2$. This cleavage is required for the protein to be able to undergo a conformational change, triggered by the low pH in endosomes, which initiates the fusion of virus and cell membranes that effects viral entry. The $HEF_1$ polypeptide

**Table 1**
Data collection.

Form I. Native ($P4_322$, $a = b = 155.4$, $c = 414.4$ Å).

| Resolution (Å) | $R_{sym}$ | Number measured | Number unique | Completeness (%) | $I > 2\sigma$ (%) |
|---|---|---|---|---|---|
| 7.27 | 0.040 | 24189 | 6560 | 0.984 | 0.992 |
| 5.89 | 0.068 | 23857 | 6379 | 0.991 | 0.972 |
| 5.18 | 0.083 | 23953 | 6316 | 0.992 | 0.959 |
| 4.72 | 0.086 | 23849 | 6292 | 0.993 | 0.955 |
| 4.39 | 0.094 | 23644 | 6232 | 0.994 | 0.949 |
| 4.14 | 0.118 | 23646 | 6253 | 0.992 | 0.924 |
| 3.93 | 0.156 | 23240 | 6187 | 0.992 | 0.885 |
| 3.77 | 0.200 | 22696 | 6183 | 0.990 | 0.848 |
| 3.62 | 0.250 | 21991 | 6200 | 0.992 | 0.793 |
| 3.50 | 0.319 | 20774 | 6160 | 0.991 | 0.731 |
| Overall | 0.100 | 231836 | 62762 | 0.991 | 0.902 |

Form II. Native ($P4_32_12$, $a = b = 217.4$, $c = 421.4$ Å).

| Resolution (Å) | $R_{sym}$ | Number measured | Number unique | Completeness (%) | $I > 2\sigma$ (%) |
|---|---|---|---|---|---|
| 8.42 | 0.026 | 27259 | 8766 | 0.984 | 0.981 |
| 6.77 | 0.033 | 27296 | 8546 | 0.993 | 0.959 |
| 5.94 | 0.043 | 27325 | 8482 | 0.993 | 0.941 |
| 5.41 | 0.051 | 27009 | 8402 | 0.992 | 0.924 |
| 5.03 | 0.049 | 26992 | 8381 | 0.990 | 0.925 |
| 4.73 | 0.049 | 26629 | 8310 | 0.988 | 0.920 |
| 4.50 | 0.053 | 26037 | 8313 | 0.987 | 0.917 |
| 4.31 | 0.057 | 25241 | 8301 | 0.987 | 0.904 |
| 4.14 | 0.064 | 24611 | 8246 | 0.987 | 0.893 |
| 4.00 | 0.074 | 24107 | 8231 | 0.988 | 0.865 |
| Overall | 0.046 | 262506 | 83978 | 0.988 | 0.923 |

Form I + 3 m*M* ethylmercury thiosalicylate (ETMS) ($P4_322$, $a = b = 155.8$, $c = 414.4$ Å).

| Resolution (Å) | $R_{sym}$ | Number measured | Number unique | Completeness (%) | $I > 2\sigma$ (%) |
|---|---|---|---|---|---|
| 6.70 | 0.091 | 19358 | 6661 | 76.4 | 99.0 |
| 5.40 | 0.100 | 25098 | 7818 | 92.5 | 97.3 |
| 4.74 | 0.097 | 24952 | 7763 | 93.1 | 97.5 |
| 4.32 | 0.100 | 24964 | 7731 | 93.2 | 97.3 |
| 4.02 | 0.109 | 24984 | 7731 | 93.5 | 95.6 |
| 3.79 | 0.121 | 25028 | 7732 | 93.9 | 93.1 |
| 3.60 | 0.139 | 24441 | 7703 | 93.9 | 90.1 |
| 3.44 | 0.163 | 23660 | 7705 | 94.1 | 85.6 |
| 3.31 | 0.198 | 22777 | 7687 | 93.6 | 79.3 |
| 3.20 | 0.253 | 21035 | 7577 | 92.5 | 72.4 |
| Overall | 0.118 | 236295 | 76108 | 91.6 | 73.5 |

**Table 1 (continued)**
Form I + 50 m*M* 9–acetomidosialic acid $\alpha$-thiomethylmercury ($P4_322$, $a = b = 154.6$, $c = 414.7$ Å).

| Resolution (Å) | $R_{sym}$ | Number measured | Completeness (%) |
|---|---|---|---|
| 12.80 | 0.072 | 1056 | 0.980 |
| 10.65 | 0.071 | 1034 | 0.987 |
| 9.46 | 0.085 | 1007 | 0.986 |
| 8.68 | 0.103 | 990 | 0.991 |
| 8.10 | 0.140 | 990 | 0.994 |
| 7.65 | 0.174 | 998 | 0.993 |
| 7.29 | 0.221 | 967 | 0.991 |
| 6.98 | 0.293 | 982 | 0.992 |
| 6.72 | 0.309 | 983 | 0.994 |
| 6.50 | 0.323 | 977 | 0.994 |
| Overall | 0.136 | 9984 | 0.990 |

Comparison with native form I.

| Resolution (Å) | $\Delta F/F$ | Number unique |
|---|---|---|
| 14.14 | 0.216 | 161 |
| 10.00 | 0.197 | 1863 |
| 8.16 | 0.188 | 2361 |
| 7.07 | 0.219 | 2748 |
| 6.50 | 0.230 | 2208 |
| Overall | 0.206 | 9341 |

contains the binding site for the 9-*O*-acetylsialic acid cell receptor and a distinct receptor-destroying, 9-*O*-acetylsialic acid esterase active site (Pleschka *et al.*, 1995). Thus, HEF contains the three activities which are distributed on two glycoproteins in influenza A and B viruses, where the HA protein has the receptor-binding and membrane-fusion activities and the neuraminidase (NA) glycoprotein has the receptor-destroying activity. Though similar in structure to the haemagglutinin (HA) of influenza A, HEF (641 residues) and HA (512 residues) share less than 15% sequence identity (Nakada *et al.*, 1984; Pfeifer & Compans, 1984). A glycoprotein on the membrane of both coronaviruses and toroviruses, the haemagglutinin-esterase (HE; ~409 residues), has 30% sequence identity to $HEF_1$ of influenza C virus and possesses the same receptor binding and esterase specificities (Luytjes et al., 1988; Vlasak *et al.*, 1988; King *et al.*, 1985). Coronaviruses cause approximately 25% of common colds in humans (Cornelissen *et al.*, 1997; Monto & Lim, 1974; McIntosh *et al.*, 1970).

Crystallization of HEF and X-ray data collection on the native HEF ectodomain from two tetragonal crystal forms [form I, $P4_{1(3)}22$, $a = b = 155.4$, $c = 414.4$ Å and form II, $P4_{1(3)}2_12$, $a = b = 217.4$, $c = 421.4$ Å] using flash cooling and synchrotron radiation have been described previously (Rosenthal *et al.*, 1996). The two crystal forms have approximately 80% solvent content and one and two HEF trimers per asymmetric unit, respectively (Rosenthal *et al.*, 1996). An enzyme inhibitor and receptor analogue which would bind in both the receptor and enzyme active sites was designed with an acetamido group substituted for the 9-*O*-acetyl group on the viral receptor 9-*O*-acetylsialic acid (Imhof *et al.*, 1988; Herrler *et al.*, 1992; Rosenthal, 1996). A mercury adduct of this inhibitor, 9-acetamidosialic acid $\alpha$-thiomethylmercuryl glycoside, was synthesized and found to form an isomorphous heavy-atom derivative (Fitz *et al.*, 1996; Rosenthal *et al.*, 1998). The structure of HEF at 3.2 Å has been described elsewhere (Rosenthal *et al.*, 1998); methods used in the structure determination are described here. Attempts to use the low-homology (<15% sequence identity) influenza A virus HA structure for molecular replacement were unsuccessful. Instead, initial SIR phases were calculated at 7.0 Å resolution from six mercury sites located by Patterson map analysis of X-ray data on the mercury derivative collected to 6.5 Å resolution. The non-crystallographic threefold-symmetry axis of the HEF trimer was defined by the mercury positions. Averaging the electron-density map calculated from SIR

phases around the molecular threefold-symmetry axis revealed an approximate molecular envelope. Iterative non-crystallographic symmetry averaging in real space, solvent flattening and histogram matching were used to extend the phases from 7.0 to 3.5 Å. The path of the polypeptide could be traced in this electron-density map and an atomic model was built. A polyalanine model was then used to solve the form II crystal containing two trimers per asymmetric unit, allowing multi-crystal averaging of the electron density. Refinement, using data to 3.2 Å, has resulted in a free $R$ factor of 26.7% and an $R_{\rm work}$ of 22.3%.

## 2. Methods and results

### 2.1. Data collection

X-ray diffraction data were collected from two crystal forms of native HEF as described previously on flash-cooled crystals at 108 K (Rosenthal *et al.*, 1996), using synchrotron radiation at the CHESS F1 and A1 beamlines and phosphor image-plate detection. 6.5 Å resolution diffraction data from a form I crystal soaked in 50 m*M* 9-acetamidosialic acid $\alpha$-thiomethylmercurylglycoside were collected with a MAR Research image-plate detector using a GX-13 rotating-anode X-ray source (Elliot Ltd.) collimated with Franks mirrors. The data were indexed and integrated with *DENZO* (Otwinowski & Minor, 1997) and further analyzed with *SCALEPACK* and the programs from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994). The highest resolution data were obtained from a form I crystal soaked in EMTS (3 m*M*) which had a small $\Delta F/F$ (0.08) similar to that found when comparing native data sets. Data-quality statistics are presented in Table 1.

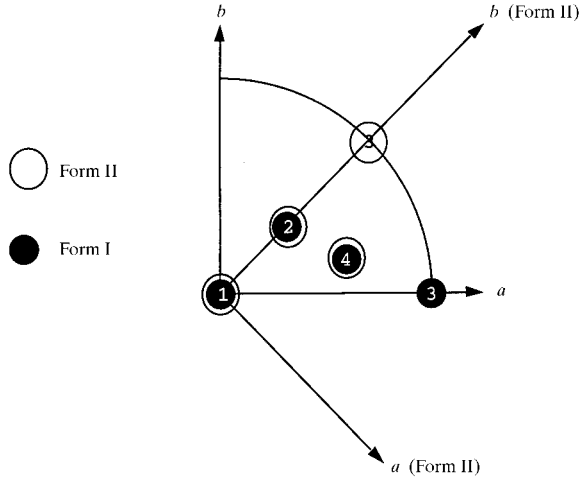



**Figure 1**
Stereographic projection of the threefold self-rotation function peaks from form I and form II crystals. Pattersons analyzed were from data in the 7.0–4.0 Å resolution annulus; vectors were restricted to the length range 25–50 Å. The lattices of crystal form I and crystal form II are related by a 45° rotation about the $c$ axis, as shown. Peaks 1, 2 and 3 are packing artifacts (see text). Peak 4 defines the molecular threefold-symmetry axis. A rotation axis is specified by $(\omega, \varphi)$ where $\omega$ is the angle between the $c$ axis and the rotation axis and $\varphi$ is the angle between the projection of the rotation axis onto the $ab$ plane and the $a$ axis.

### 2.2. Self-rotation function and molecular-replacement calculations

The orientation of the non-crystallographic threefold-symmetry axis in the HEF crystals was identified by analysis of self-rotation functions, but locating the HEF molecule with cross-rotation and translation functions using the influenza A HA structure was unsuccessful. Searches for the non-crystallographic molecular threefold axis were performed using both reciprocal- and real-space methods. [Model calculations using the influenza A HA structure placed in different orientations in a model $P4_{1(3)}22$ unit cell indicated that the correct threefold self-rotation function peak could be found in model data using the *X-PLOR* (Brünger, 1992*b*) real-space rotation function searching with a 25–50 Å vector annulus of the Patterson.] Fig. 1 shows the top four *X-PLOR* self-rotation

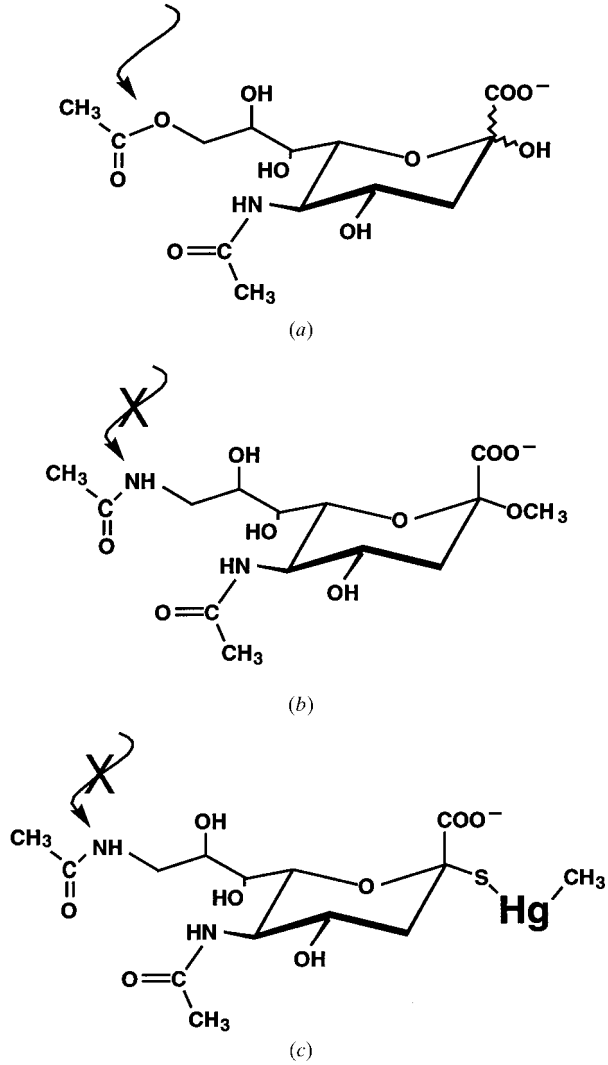



**Figure 2**
Sialosides which bind HEF. (*a*) HEF cell receptor 9-*O*-acetyl-sialic acid. The HEF esterase removes the 9-*O*-acetyl group (arrow). (*b*) A non-hydrolyzable substrate analog and enzyme inhibitor, 9-acetamidosialic acid $\alpha$-methyl glycoside. (*c*) The Hg derivative, 9-acetamidosialic acid $\alpha$-thiomethylmercuryl glycoside, is a cell-receptor analog as well as a enzyme inhibitor. It binds to both the cell receptor-binding site and the enzyme active site.

function peaks for HEF crystal forms I and II. The unit cells of these crystal forms are related by a 45° rotation about their common *c* axes, as described previously (Rosenthal *et al.*, 1996). Peaks 1, 2 and 3 in each space group appear to be a consequence of the superposition of crystal packing vectors, as they were also found as the top peaks using the observed

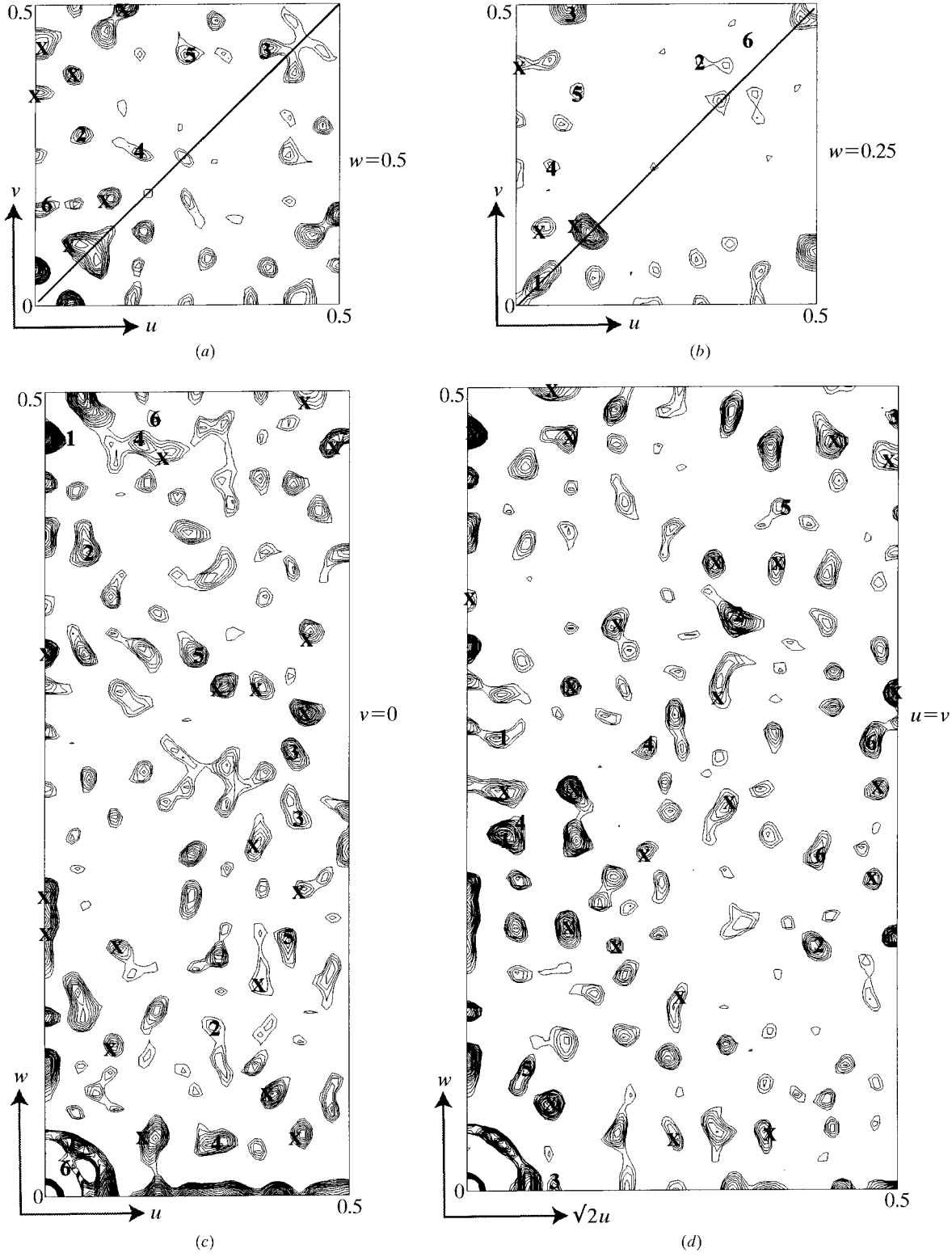



**Figure 3**
Harker sections in the mercury difference Patterson maps. The mercury compound is described in Fig. 2(*b*). The Patterson was calculated with 15–6.5 Å data. Peaks on sections $w = \frac{1}{2}$ and $w = \frac{1}{4}$ are related by mirror symmetry across the diagonal. The asymmetric units of the $v = 0$ and $u = v$ sections are shown. The Harker peaks from the six mercury positions summarized in Table 2 are indicated by the corresponding numbers. Cross vectors close to the Harker sections are marked as 'X'.

intensities randomized with respect to Miller indices (*hkl*) within resolution shells, thus preserving the intensity profile as a function of resolution similar to the original data (Jones, Walker *et al.*, 1991). This suggested that peaks 1–3 were rotation-function artifacts and that peak 4, common to both space groups, identified the orientations of the molecular threefold-symmetry axes ($\omega$ = 63, $\varphi$ = 20°; Fig. 1 angle conventions). The subsequent determination of the location of heavy-atom derivatives labeling the three receptor-binding and three enzyme active sites confirmed this orientation of the non-crystallographic symmetry axis.

HEF has been expected to resemble the influenza A virus HA, for which the crystal structure is known (Wilson *et al.*, 1981; Wiley *et al.*, 1981; Weis *et al.*, 1990), for the following reasons.

(i) HEF and HA share the activities of binding a sialic acid and mediating membrane fusion.

(ii) Low-resolution image reconstructions (∼30 Å) from electron micrographs of the hexagonal glycoprotein arrays on influenza C virus in negative stain (Hewat *et al.*, 1984) suggested that HEF is a trimer similar in size and shape to the 135 Å long and 50 Å wide influenza A haemagglutinin (HA; Wilson *et al.*, 1981).

(iii) Although the overall sequence identity is less than 15% and attempts to align the $HA_1$ and $HEF_1$ sequences in the absence of the structure of HEF were unsuccessful (Pfeifer & Compans, 1984; Nakada *et al.*, 1984), $HEF_2$ and $HA_2$ are the same length and share 15% sequence identity (Pfeifer & Compans, 1984; Nakada *et al.*, 1984).

The similarity of $HEF_2$ to $HA_2$ includes three conserved cysteines, a similar non-polar fusion-peptide at the N-terminus and a membrane-anchor sequence near the C-terminus. The sequence of $HEF_1$ is 100 residues longer than $HA_1$. The extra 100 residues have been expected to form the esterase domain unique to $HEF_1$, but they could not be located in the $HEF_1$ sequence by sequence comparisons.

Cross-rotation functions were calculated using both reciprocal- and real-space methods with the influenza A virus HA trimer as a search model or subsets of the model corresponding to the HA monomer, the $HA_1$ subunits only, the $HA_2$ subunits only and the same models consisting of only $C^{\alpha}$

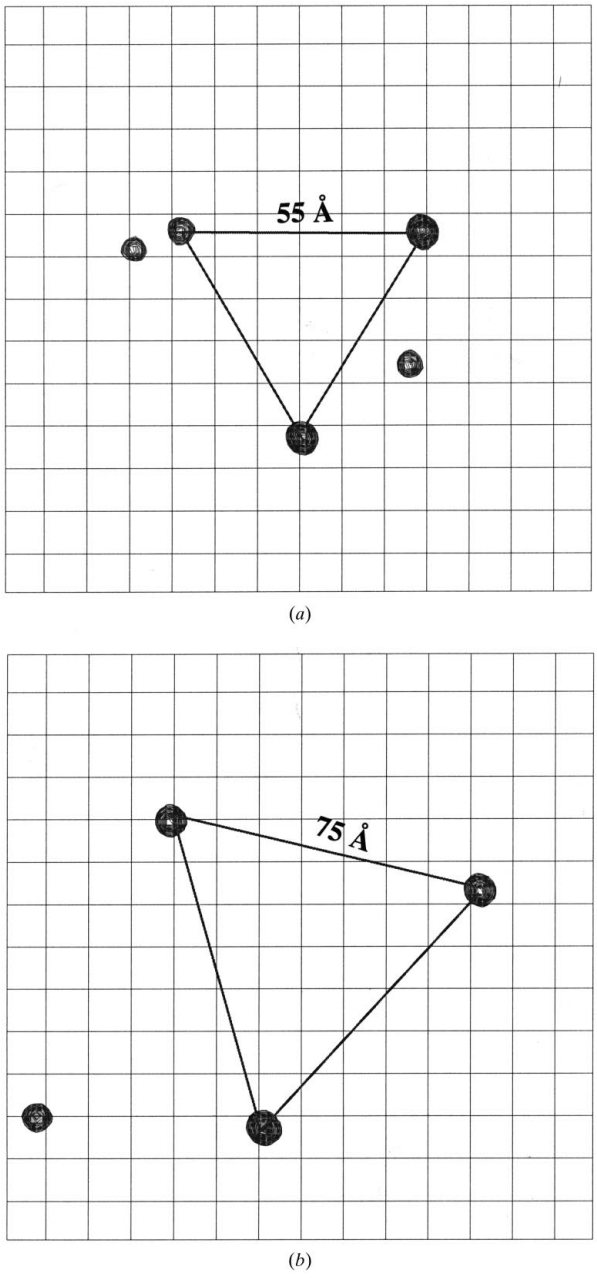



**Figure 4**
Skewed difference Fourier map (between the derivative and the native data) using phases from the three mercury sites with highest occupancies (Table 2). The map is a view down the molecular threefold-symmetry axis which was defined by the location of the three heavy atoms forming a triangle in Fig. 4(*b*). (*a*) Three Hg atoms located at the HEF receptor-binding sites. (*b*) Three Hg atoms located at the HEF esterase sites (see text). The other nearby peaks in the diagram are related by crystallographic symmetry.

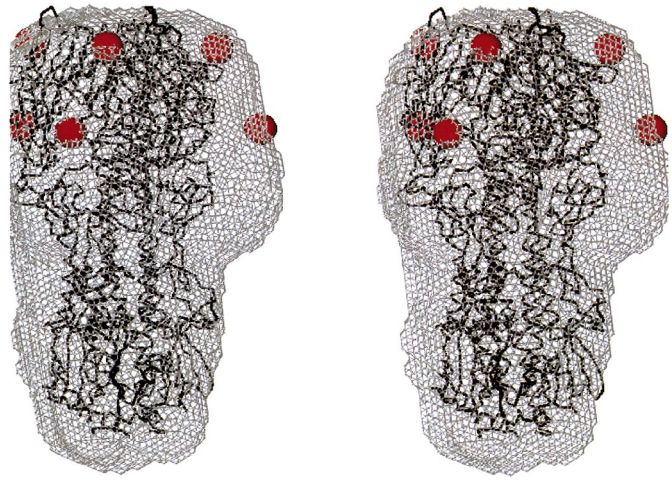


**Figure 5**
Stereographic drawing of the initial molecular envelope of the HEF trimer, superimposed on the HA $C^{\alpha}$ model. The Hg-atom positions are marked as red spheres. The HA trimer from influenza A virus was positioned such that its receptor-binding sites are near the Hg atoms forming the smaller triangle. The other Hg atoms are expected to be near the enzyme active sites. The bulge in the envelope relatively near the lower triangle of Hg atoms defines the location of the enzyme domain of HEF.

atoms. No consistent peaks were found for different models, data-resolution ranges (including the low-resolution 15–8 Å annulus, see below) and Patterson search annuli. Patterson correlation refinement (Brünger, 1990) and translation functions calculated in both $P4_122$ and $P4_322$ using the top 100 peaks for all the rotation searches failed to give a translation-function peak above the noise level.

In retrospect, the 9-acetamidosialic acid $\alpha$-thiomethylmercurylglycoside heavy-atom positions confirmed the self-rotation function analysis. Although several cross-rotation function peaks were near the self-rotation peak, no significant translation-function peaks were found. To investigate whether searches with lower resolution data (not actually measured) might have been successful, low-resolution data were calculated from the final model and solvent correction added. A real-space cross-rotation function calculation with model data in the resolution range 40–10 Å using *X-PLOR* (Brünger, 1992*b*) contained the approximately correct orientation within the top ten peaks (no solution was evident in 100–10 Å resolution data). However, no solution closer than 30 Å to the correct location was found by translation-function searches with *X-PLOR* (Brünger, 1992*b*). To investigate further why the molecular-replacement searches with the HA structure were unsuccessful, the HA structure was placed at the location of the final HEF structure and refined against the HEF data, allowing each polypeptide to move as a rigid body. This refined model has a correlation coefficient of 0.22 and an *R* factor of 53% against the HEF data in the resolution range 15.0–3.5 Å. The mean phase error between the refined HEF structure and the HA model is 89°. When the HA model was placed at incorrect locations in the form I unit cell, the correlation coefficient was 0.20 and the *R* factor was 55%. The structural differences between HEF and HA are significant. Although the structure of individual sequence segments of the two proteins can be individually superimposed quite closely (Rosenthal *et al.*, 1998), when the full structures are compared, inter-domain movements over the length of the large molecule contribute to r.m.s. discrepancies between $C^{\alpha}$ positions of 11 Å between $HEF_1$ and $HA_1$ and 9 Å between $HEF_2$ and $HA_2$, excluding HEF segments not present in HA. These large structural changes are apparently responsible for the failure of the molecular-replacement searches.

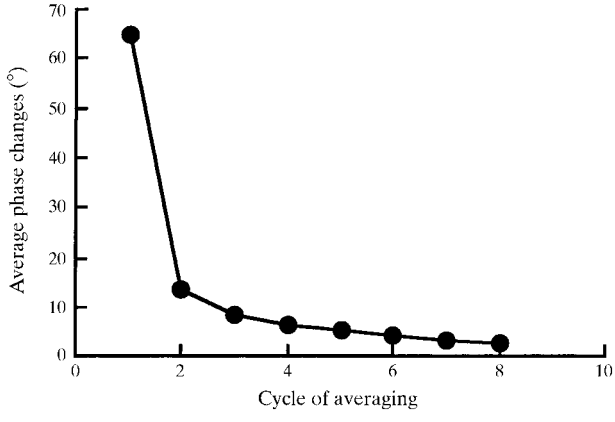



**Figure 6**
Average phase changes per cycle of iterative threefold real-space averaging at 7.0 Å resolution. Convergence was achieved after five or six cycles.

### 2.3. Design of an HEF-specific heavy-atom derivative

HEF contains a receptor-binding site for 9-*O*-acetylsialic acid, the virus receptor, and a separate 9-*O*-acetylesterase enzyme active site, which removes the 9-*O*-acetyl group as depicted in Fig. 2(*a*) (Herrler *et al.*, 1988; Vlasak *et al.*, 1989;

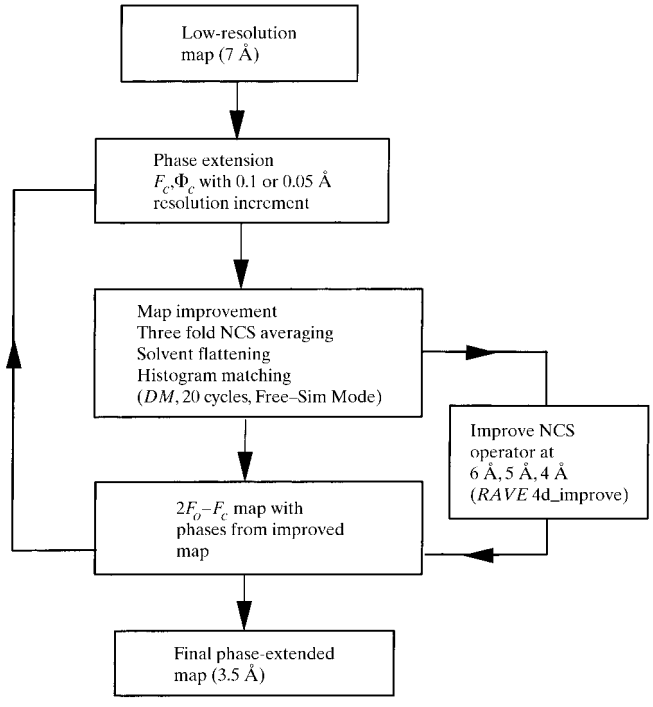



**Figure 7**
Flow diagram of the phase-extension procedure from 7.0 to 3.5 Å resolution.

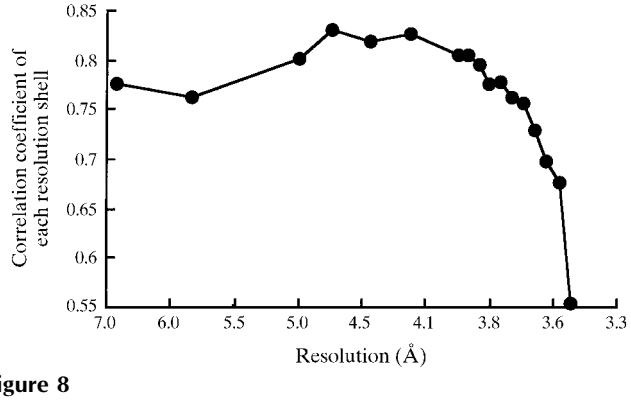
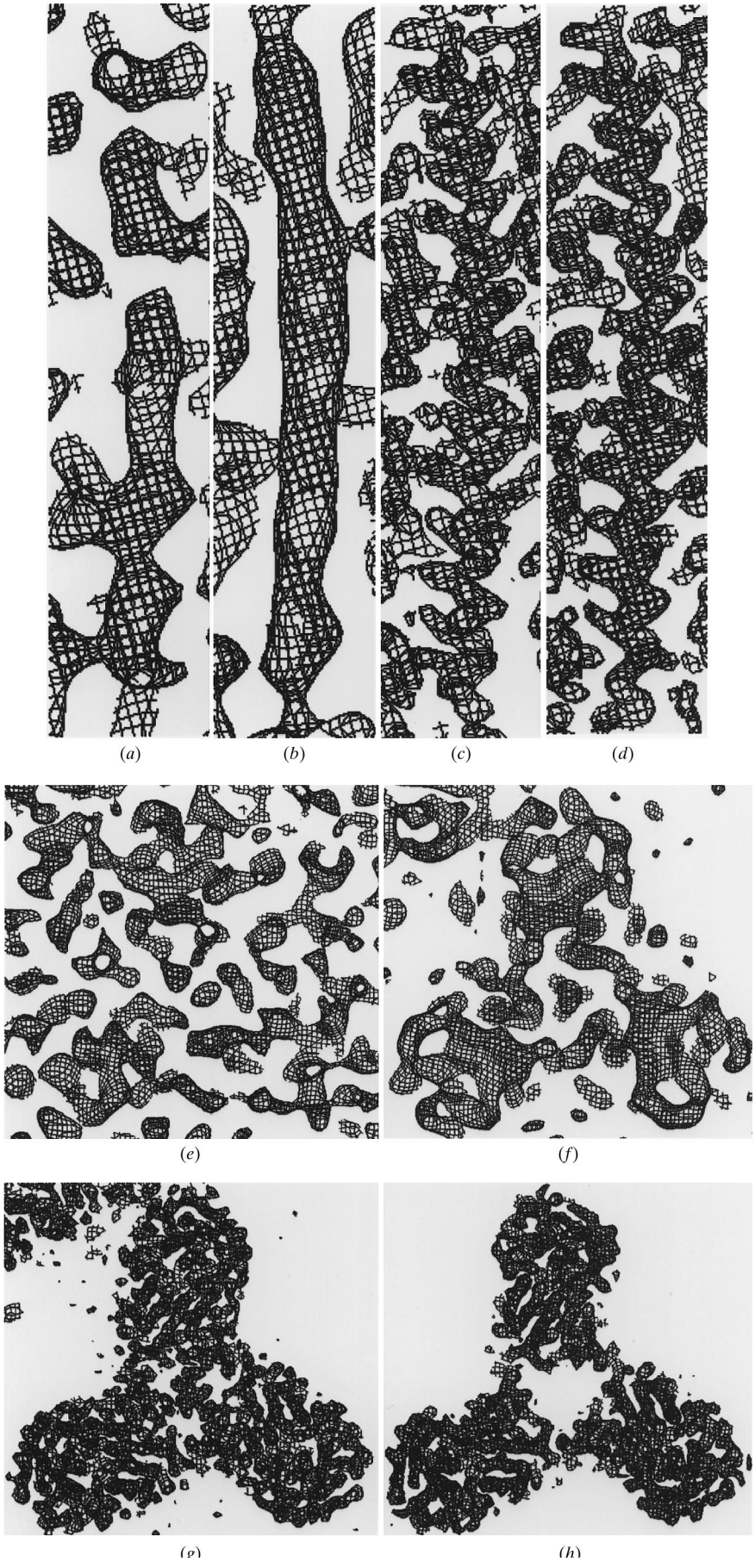



**Figure 8**
Correlation coefficient of each resolution shell on the final phase-extended electron-density map at 3.5 Å resolution. The correlation coefficient between the structure factors calculated from the phase-extended electron-density map and the observed structure factors is approximately constant to 3.8 Å, after which it decreases. In the highest resolution shell (3.6–3.5 Å), the correlation coefficient is still above 0.5. The overall correlation coefficient of the 3.5 Å resolution phase-extended map is 0.79.




Hayes & Varki, 1989; Pleschka *et al.*, 1995). The structural study of HEF in complex with its natural receptor is hindered by the presence of the receptor-destroying function of the HEF esterase. A non-hydrolyzable substrate analogue and enzyme inhibitor, 9-acetamidosialic acid $\alpha$-methylglycoside, is shown in Fig. 2(*b*) and has a $K_i$ of 2.8 m*M* (Rosenthal, 1996; Imhof *et al.*, 1988). Because this inhibitor also prevents haemagglutination (Fitz *et al.*, 1996), it was expected that it might bind to the receptor-binding site as well as to the enzyme active site.

A mercury adduct of the inhibitor, 9-acetamidosialic acid $\alpha$-thiomethylmercuryl glycoside (Fig. 2*c*), was synthesized with the mercury at the glycosidic linkage point where the natural receptor is linked to a cell-surface oligosaccharide and, therefore, unlikely to influence binding to HEF. This mercury-substituted sialoside was found to be a competitive inhibitor of enzyme activity with a $K_i$ = 4.2 m*M* (Fitz *et al.*, 1996). Although this is a weak inhibition constant, HEF crystals could be soaked in 50 m*M* ($10 \times K_D$) solutions of inhibitor to achieve high occupancy.

### 2.4. Locating the mercury positions

X-ray diffraction data collected to 6.5 Å resolution from a form I crystal soaked in the 9-acetamidosialic acid $\alpha$-thiomethylmercurylglycoside indicated that the derivative had similar unit-cell constants to the native crystal and a mean isomorphous difference in structure factors

**Figure 9**
Electron density compared at different stages in the phase-improvement and phase-extension procedures (contoured at 1$\sigma$.) (*a*) SIR electron density at 7.0 Å along the long $HEF_2$ helix. (*b*) Electron density of the same section as in (*a*) after improvement by iterative cycles of threefold NCS averaging, solvent flattening and histogram matching. (*c*) 3.5 Å phase-extended electron density of the same section as in (*a*). (*d*) Final refined electron density at 3.2 Å resolution. (*e*) SIR electron density at 7.0 Å resolution viewed from the top of the molecule down the molecular threefold axis. (*f*) 7.0 Å phase-improved electron density showing the same section as in (*e*). The protein density is clearly separated from the solvent density and three individual monomers are distinguishable. $\beta$-strands appear as continuous sheets. (*g*) 3.5 Å resolution phase-extended electron density of the same section as in (*e*). Individual $\beta$-strands are distinguishable. (*h*) Final 3.2 Å refined electron density.

**Table 2**
Heavy-atom parameters and SIR phasing statistics.

| Sites | Positions | | | Occupancy |
|---|---|---|---|---|
| | *X* (Å) | *Y* (Å) | *Z* (Å) | |
| 1 | 81.1 | 76.2 | 6.2 | 1.02 |
| 2 | 99.2 | 147.9 | 21.1 | 0.95 |
| 3 | 44.2 | 124.2 | 49.0 | 0.91 |
| 4 | 56.2 | 89.8 | 6.1 | 0.91 |
| 5 | 45.3 | 96.5 | 33.9 | 0.87 |
| 6 | 74.8 | 143.1 | 5.2 | 0.65 |

| | |
|---|---|
| Resolution (Å) | 15.0–7.0 |
| Total reflections | 7088 |
| Phasing power† | |
| Centric | 0.90 |
| Acentric | 1.29 |
| $R_{\rm Cullis}$‡ | |
| Centric | 0.76 |
| Acentric | 0.85 |
| Mean FOM | |
| Centric | 0.402 |
| Acentric | 0.235 |
| Overall | 0.264 |

† Phasing power = $\langle F_H/(\lvert F_{PH} - \lvert F_P + F_H\rvert\rvert)\rangle$. ‡ $R_{\rm Cullis}$ = $(\lvert F_{PHi} - \lvert F_P + F_H\rvert\rvert)/(\lvert F_{PHi} - F_P\rvert)$.

($\Delta F/F$) of 20.6% (Table 1). The difference Patterson map calculated at 6.5 Å resolution was noisy (Fig. 3). In addition to factors such as non-isomorphism, we anticipated a poor signal-to-noise ratio for the mercury difference Patterson owing to the large molecular mass of protein in the asymmetric unit (225 kDa). The six mercury positions were located by a combination of manual and computational searches of the Patterson map. Beginning with some prominent peaks on the Harker sections, a few tentative sites were assigned, which were confirmed by searches for the cross vectors among the sites using the Patterson search program *RSPS* (Collaborative Computational Project, Number 4, 1994). A combination of further Patterson searches and difference Fourier calculations located the full six expected sites. The location of the Harker vectors of the six mercury sites are marked on the four Harker sections (Fig. 3). Cross vectors which overlap with or are close to the Harker sections are marked with an 'X' in Fig. 3. Distances between the sites indicated that they formed two equilateral triangles, as anticipated for the two sialoside-binding sites on each monomer of the trimeric HEF. The difference Fourier electron-density maps skewed for viewing normal to the planes of the triangles are shown in Fig. 4. The molecular threefold-symmetry axis defined by the mercury locations is oriented $\omega = 63^\circ$ from the *c* axis and $\psi = -20^\circ$ from the *a* axis, according to the angle conventions of Fig. 1, at the same place indicated by the self-rotation function analysis. The positions of the mercury sites are listed in Table 2.

The three mercury sites 55 Å from each other (Fig. 4*a*) form a triangle similar in size to that formed by the locations of the receptor-binding sites on the influenza A virus HA trimer. When the receptor sites of the HA were positioned near these mercury positions, the second triangle of mercury sites 75 Å from each other (Fig. 4*b*) fell into space off the edge of the HA model, suggesting the location of the enzyme domain of HEF (Fig. 5). The positions of the heavy-atom sites were refined and SIR phases were calculated with *MLPHARE* (Otwinowski & Minor, 1997). Statistics are summarized in Table 2.

### 2.5. Molecular-envelope determination and phase improvement at 7.0 Å resolution

An SIR electron-density map was calculated to 7.0 Å resolution using the six mercury sites. The map revealed long rod-like features reminiscent of the long triple-stranded $\alpha$-helical coiled coil in $HA_2$. The atomic model of the HA was placed along the trimer axis defined by the heavy-atom positions to facilitate the calculation of approximate transformation matrices which superimposed the electron density of monomers onto each other in the HEF trimer using *O* (Jones, Zou *et al.*, 1991). These approximate non-crystallographic symmetry operators were improved by a real-space six-dimensional search in order to find the maximum correlation between electron densities related by the non-crystallographic symmetry axis using *RAVE* (Kleywegt & Jones, 1994).

To permit iterative real-space averaging, solvent flattening and histogram matching, an initial molecular envelope was constructed from the SIR electron-density map at 7.0 Å resolution, assuming 78% solvent content. An SIR electron-density map was threefold averaged about the molecular symmetry axis without an envelope and folded back into the asymmetric unit using *RAVE* (Kleywegt & Jones, 1994). This map was back-transformed to calculate structures factors which were input to the program *DM* for determining the solvent mask in a unit cell (Cowtan, 1994). The mask about one molecule was isolated by averaging the *DM* solvent mask around a molecular threefold axis using *RAVE* (Kleywegt & Jones, 1994). Positions in the *DM* mask not reinforced by the local threefold averaging were set to zero with *MAPMAN* (Jones, Zou *et al.*, 1991). The mask was enlarged and smoothed by five-point interpolation at its boundaries (*MAMA*; Kleywegt & Jones, 1996*a*,*b*). Internal 'islands' were removed and crystallographic packing overlaps minimized using *MAMA*. The molecular envelope (Fig. 5) was approximately the size and shape of the influenza HA, but contained a new domain in the vicinity of the second triangle of mercury sites, formed by the additional 100 residues of $HEF_1$.

With the molecular envelope and the NCS transformation matrices, iterative threefold real-space averaging with solvent flattening was used to improve the phases to 7.0 Å resolution using *RAVE* (Kleywegt & Jones, 1994). Convergence was observed after five or six cycles, as indicated by the average phase changes between consecutive cycles dropping from 65° to less than 3° (Fig. 6). The increase in the quality of the electron-density map at 7.0 Å resolution was dramatic, with continuous rods of $\alpha$-helical density replacing broken rod-like fragments in the stem region of the molecule (Figs. 9*a*, 9*b*) and connected electron density clearly defining the top globular domains (Figs. 9*e*, 9*f*).

### 2.6. Phase extension from 7.0 to 3.5 Å resolution

The phases of the improved electron-density map at 7.0 Å resolution were then extended to 3.5 Å resolution by the procedure outlined in Fig. 7. At each resolution step, the electron-density map was back-transformed to calculate phases for slightly higher resolution Fourier terms than were used to calculate the map. Electron-density maps calculated with the incrementally higher resolution terms were improved by 20 cycles of threefold NCS averaging in real space, solvent flattening and histogram matching using *DM* (Cowtan, 1994). $\sigma_A$ weighting (Free-Sim mode in *DM*) was used to combine phases between the iterative cycles. A total of 50 phase-extension steps were used, which included 20 0.1 Å steps from 7.0 to 5.0 Å and 30 0.05 Å steps from 5.0 to 3.5 Å (each resolution increment was less than the interval between lattice points along the longest unit-cell axis). At resolutions of 6.0, 5.0 and 4.0 Å, the NCS operators were redetermined using a six-dimensional search for the maximum correlation NCS-related electron density. The progress of the phase-extension process was monitored by the *DM* free-*R* indicator and the correlation coefficient between the observed amplitudes and those calculated from the map (Table 3). The final phase-extended electron density calculated at 3.5 Å had good correlation coefficients in all resolution shells (Fig. 8) and showed clearly interpretable protein features including long $\alpha$-helices in the stem region (Fig. 9*c*) and a $\beta$-barrel in the top globular domain (Fig. 9*g*). The hand of the $\alpha$-helices was used to determine the correct enantiomorph of the space group to be $P4_322$.

**Table 3**
*R* factors and correlation coefficients of phase-extended maps at different resolutions during the phase-extension procedure.

| Resolution limit (Å) | *R* factor (overall) | Correlation coefficient (overall) |
|---|---|---|
| 7.00 | 0.373 | 0.770 |
| 6.00 | 0.352 | 0.771 |
| 5.00 | 0.279 | 0.791 |
| 4.75 | 0.271 | 0.797 |
| 4.50 | 0.271 | 0.794 |
| 4.25 | 0.264 | 0.807 |
| 4.00 | 0.264 | 0.805 |
| 3.75 | 0.264 | 0.800 |
| 3.50 | 0.271 | 0.786 |

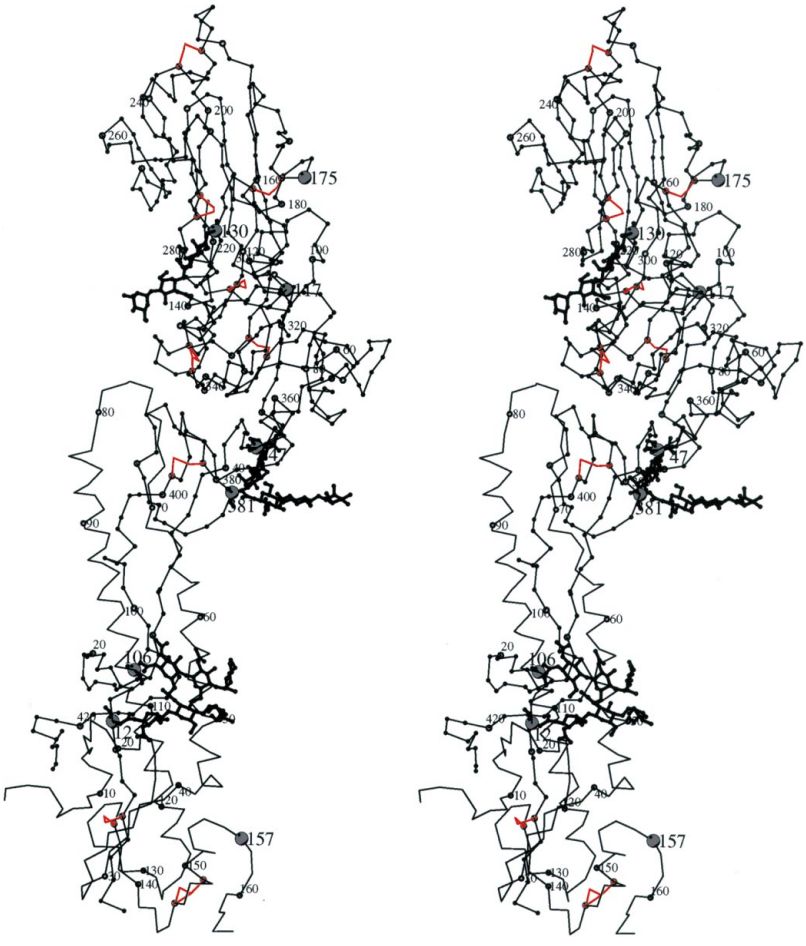


**Figure 10**
Stereographic diagram of the $C^{\alpha}$ path of the HEF monomer. [Every 20 residues in $HEF_1$ (open circles) and every ten residues in $HEF_2$ (open circles) are numbered.] The nine disulfide bridges are in red. Glycosylation sites are marked by large gray circles. Carbohydrate at five sites is shown in bold.

### 2.7. Atomic model built into the 3.5 Å resolution electron density

97% of the path of the $HEF_1$ chain and 85% of the path of the $HEF_2$ chain were traced into the phase-extended electron-density map at 3.5 Å resolution. A partial molecular model for $HEF_1$, including residues 4–424 ($HEF_1$ has 432 residues), was built using the amino-acid sequence aided by clear electron density for some of the aromatic side chains, seven internal disulfide bridges and N-linked carbohydrate. Only 131 of the 175 residues of the $HEF_2$ polypeptide could be built unambiguously (residues 24–154). However, additional electron density at the N- and C-termini of $HEF_2$ suggested where the remaining residues might be located. The phase-extension protocol, as described above, was repeated with a new envelope calculated from the partial model of $HEF_1$ and $HEF_2$ (Madden, 1992) but extended to include the untraced electron density. The path of $HEF_2$ was indicated for residues 8–154, though side-chain density was poor.

Further clarity in the chain tracing came from an analysis of the form II crystal. Diffraction data collected to 4.0 Å resolution from crystal form II, which contains two HEF trimers per asymmetric unit, permitted iterative ninefold averaging to 4.0 Å resolution of electron-density maps from form I and form II crystals. The location of the HEF trimers in the form II crystals was determined by using a polyalanine model of the unrefined HEF trimer and the program *AMoRe* (Navaza, 1994). A cross-

rotation function indicated trimers at the same orientation as in crystal form I but rotated 45° about the *c* axis, as expected (Rosenthal *et al.*, 1996). The top rotation-function peak (8.7$\sigma$) gave a translation-function solution in $P4_32_12$ with a correlation coefficient of 0.23 (18.9$\sigma$). With the first trimer fixed, the second trimer was located in a translation function (49.5$\sigma$) at a relative translation vector of (0.481, 0.498, 0.005) in fractional coordinates. This displacement is a pseudo-*C*-centering translation ($\frac{1}{2}$, $\frac{1}{2}$, 0) between the two trimers in the asymmetric unit, consistent with a peak in the native Patterson function of the form II crystals (Rosenthal *et al.*, 1996). Rigid-body refinement of both trimers yielded a correlation coefficient of 0.38 and an *R* factor of 0.48. Side-chain electron densities consistent with the HEF sequence were evident in a $2F_o - F_c$ map calculated from the polyalanine model.

A model of the two trimers containing $HEF_1$ residues 6–420 and $HEF_2$ residues 30–154, which were the clearest residues in the electron-density maps of the form I crystal before refinement, was used to calculate a $2F_o - F_c$ map in the form II crystal to 4.0 Å resolution. Structure factors calculated from this electron-density map were used with the program *DM* for 20 cycles of iterative sixfold non-crystallographic symmetry averaging in real space, solvent flattening and histogram matching. The molecular envelopes for this procedure were calculated from the model from the form I crystal with *ENVAT* (Madden, 1992), extended in the regions where the original model was expected to be incomplete with *MAMA* (Kleywegt & Jones, 1996*b*), and trimmed to avoid symmetry overlaps using the program *NCSMASK* (Collaborative Computational Project, Number 4, 1994). A monomer mask was used for non-crystallographic symmetry averaging and a mask surrounding both trimers was used for solvent flattening. The improved sixfold-averaged electron-density map from the form II crystal was combined with the threefold-averaged electron-density map calculated prior to model refinement from the form I crystals in 20 cycles of iterative ninefold averaging in real space, solvent flattening and histogram matching of monomer density in both crystal forms using the program *DMmulti* (Cowtan, 1994). The resulting electron-density map, although at lower resolution (4.0 Å) than those from form I alone (3.2 Å), showed connected electron density for $HEF_1$ residues 3–425 and $HEF_2$ residues 6–154. The improved electron density for $HEF_2$ residues 6–24 resolved uncertainty about the path of main chain in this region, which had been confused by the presence of the electron density from an oligosaccharide attached at $HEF_1$ residue 12.

### 2.8. Refinement and model rebuilding

The initial model had an *R* factor of 42% for all the intensities in the 15.0–3.5 Å resolution range. 5% of the diffraction data (3700 reflections) were omitted from subse-

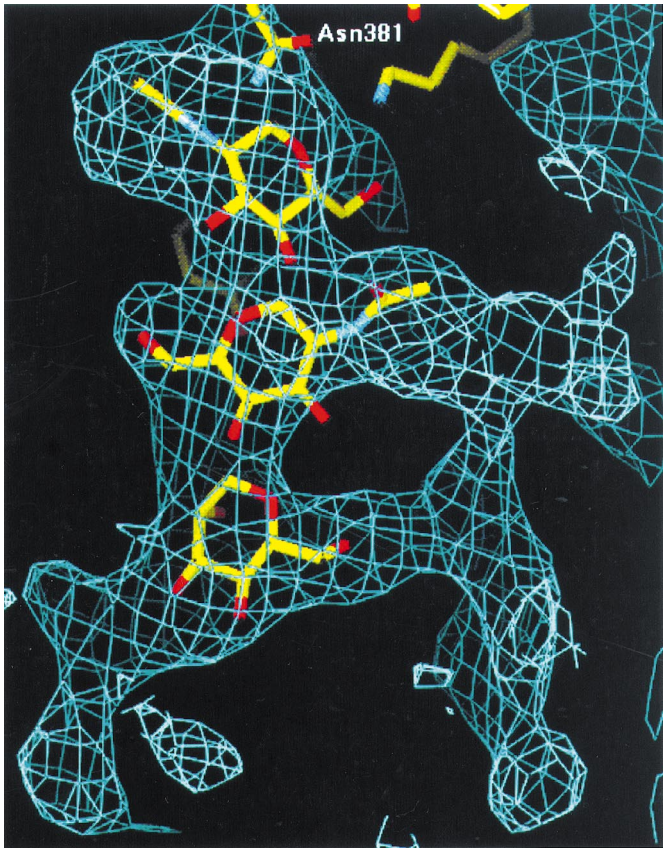



**Figure 11**
Electron density and superimposed carbohydrate model at Asn381. The electron density was calculated by omitting the carbohydrate model (contoured at 1$\sigma$).

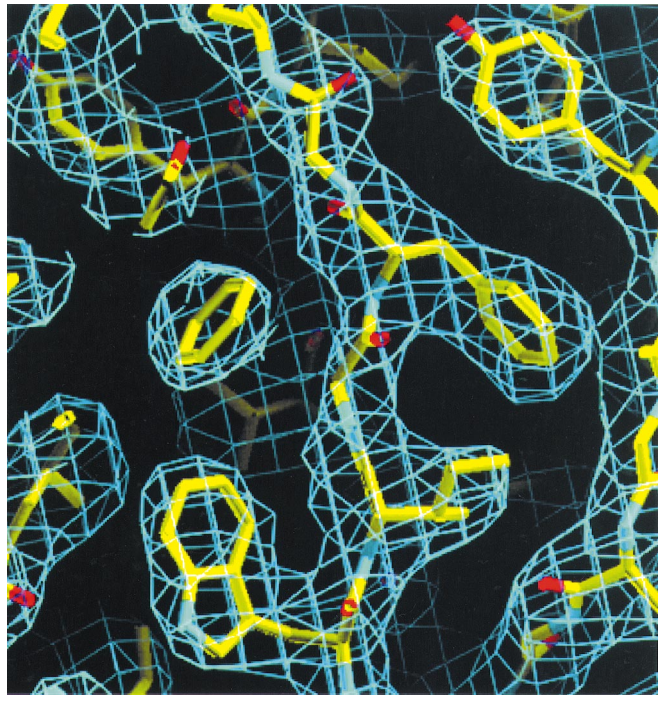


**Figure 12**
Final refined model at 3.2 Å resolution superimposed on a simulated-annealing omit map. The electron-density map was calculated by omitting atoms within an 11 Å radius surrounding residue 53 of $HEF_1$, with an initial annealing temperature of 1000 K (Brünger, 1992*b*; Hodel *et al.*, 1992). The map was calculated with data between 15 and 3.2 Å resolution and contoured at 1$\sigma$.

quent refinement and used to calculate the free $R$ factor (Brünger, 1992*a*). Strict non-crystallographic threefold symmetry was applied in the initial rounds of refinement (Weis *et al.*, 1990; Braig *et al.*, 1995; Brünger, 1992*b*), with an observations-to-parameters ratio of 4.1 (60000 reflections, 14630 atoms). An initial $B$ factor of 40 Å$^2$ was assigned to every atom; Wilson statistics suggested an overall $B$ factor of 60 Å$^2$ based on observed diffraction intensities.

An initial 150 cycles of positional refinement decreased $R_{\rm free}$ to 36.1% and $R_{\rm work}$ to 34.2% (*X-PLOR*; Brünger, 1992*b*; Table 4). This was followed by simulated-annealing refinement with slow cooling, during which the temperature was decreased from 4000 to 300 K by increments of 25 K every 12.5 fs. 120 cycles of conventional energy minimization were performed after the simulated annealing. The $R$ factors decreased more than 2% to an $R_{\rm free}$ of 33.7% and an $R_{\rm work}$ of 31.2% (Table 4). A restrained temperature ($B$) factor refinement which restricts the deviation of $B$ factors between adjacent atoms to less than $2\sigma$ was applied after simulated annealing (Table 4). The Ramachandran plot was consulted during refinement using *PROCHECK* (Laskowski *et al.*, 1993) and residues outside the most probable regions were re-examined and rebuilt if consistent with the electron density.

The phases calculated from the refined model were used to calculate a $\sigma_A$-weighted electron-density map using ($2mF_o - DF_c$) coefficients (Read, 1986; Kleywegt & Jones, 1996*a*,*b*). This electron-density map was threefold averaged and solvent flattened using *RAVE* and an envelope calculated from the model (*ENVAT*; Madden, 1992). There was a clear improvement in the electron density as a result of the

**Table 4**
Refinement statistics.

80513 reflections for $R_{\rm work}$, 4316 for $R_{\rm free}$. Resolution 10.0–3.53 Å unless otherwise noted.

| Model | $R_{\rm work}$ (%) | $R_{\rm free}$ (%) |
|---|---|---|
| Initial model (A4–A424, B24–B154) | 42 | 42 |
| Positional refinement (*X-PLOR*) | 34.2 | 36.1 |
| Simulated-annealing refinement (*X-PLOR*) | 31.2 | 33.7 |
| Restrained $B$-factor refinement | 27.0 | 30.7 |
| Building and rebuilding (A1–A428, B6–B165) | 22.7 | 27.9 |
| Add CHO | 21.7 | 27.2 |
| 10–3.2 Å resolution | | |
| Overall | 22.3 | 26.7 |
| Last resolution shell | 30.5 | 35.3 |

| | |
|---|---|
| R.m.s. deviations | |
| From ideal bond length (Å) | 0.014 |
| From ideal bond angle (°) | 1.83 |
| Average $B$ factors (Å$^2$) | |
| Main chain | 55.2 ± 2.2 |
| Side chain | 58.3 ± 4.2 |

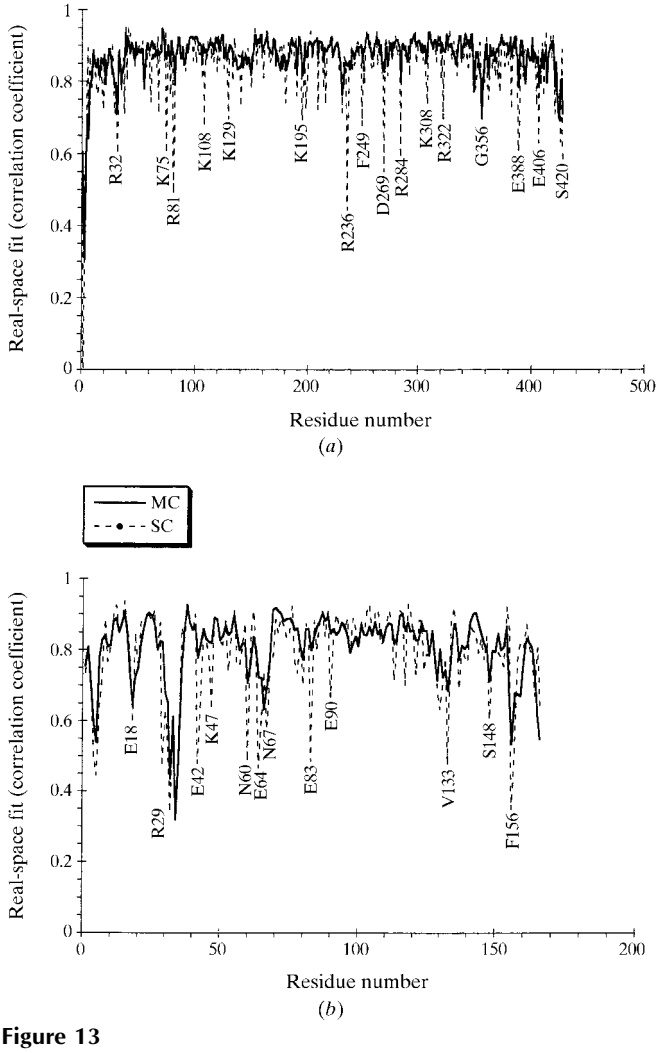



**Figure 13**
Correlation coefficient indicating real-space fit between the $2F_o - F_c$ electron density and the final model refined at 3.2 Å resolution for main-chain atoms (MC, solid line) or side-chain atoms (SC, dashed line). Solvent-exposed polar side chains are labeled. (*a*) HEF$_1$, (*b*) HEF$_2$.

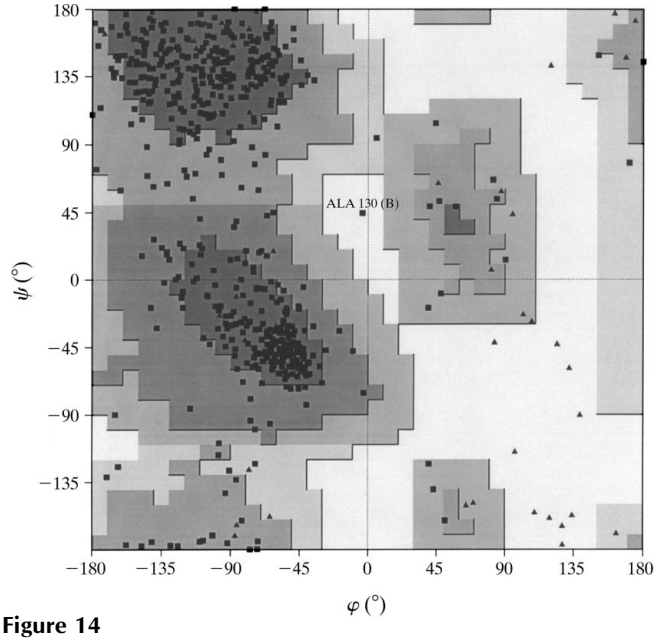



**Figure 14**
Ramachandran plot of the main-chain torsion angles ($\varphi$, $\psi$) edited from *PROCHECK* for the final HEF model refined at 3.2 Å resolution. 76% of the residues have dihedral angles in the most favored region (darkest gray) and 0.2% of the residues are in the disallowed region (white). Triangles represent glycines and squares represent all other residues.

refinement, including better electron density for $HEF_1$ residues 1–3 and 425–427 and $HEF_2$ residues 8–23, 155–160, and confirmation of the path of the polypeptide chain segments which had been omitted from the refinement.

These additional residues were added to the model and refined together with the rest of the molecule as described above. Electron density for residues 1–5 and 161–175 of $HEF_2$ and residues 427–432 of $HEF_1$, all at termini, remained poor after many cycles of refinement and examination of averaged difference Fourier maps. There are crystal contacts near the $HEF_1$ and $HEF_2$ C-termini which can cause deviations from the non-crystallographic symmetry. A tentative model including residues 1–8 and 161–175 of $HEF_2$ and 427–432 of $HEF_1$, was built into unaveraged $2mF_o - DF_c$ Fourier maps. This complete, but tentative, model was refined by applying a weighted non-crystallographic restraint (*X-PLOR*; Brünger, 1992*b*) for the entire model and residues 422–432 of $HEF_1$, 1–10 and 161–175 of $HEF_2$ free of any restraint. No dramatic changes were observed in the unrestrained residues after refinement.

In each round of refinement, before simulated annealing, *B* factors were first refined as a group with each residue given one *B* factor for all the main-chain atoms and one for all the side-chain atoms. A restraint *B*-factor refinement which restricts the deviation of the *B* factors between adjacent atoms to be less than $2\sigma$ was applied after simulated-annealing refinement.

There are eight potential N-linked carbohydrate attachment sites per HEF monomer, located at $HEF_1$ residues 12, 47, 117, 130, 176 and 381 and $HEF_2$ residues 106 and 157 (Fig. 10). At each of five of the sites ($HEF_1$ 12, 47, 130, 381, and $HEF_2$ 106), clear electron density was observed for a trisaccharide attached to the asparagine (*e.g.* Fig. 11). A model of MAN–NAG–NAG was built into the electron density for each of these sites (Fig. 11).

Refinement of this addition to the model resulted in very small changes in the *R* factors (Table 4). Extra electron density was observed beyond the core trisaccharide at residues 47, 130 and 381. Some density was observed near residue 176 in one monomer, but no further carbohydrate could be added to the model with confidence. Of the two potential oligosaccharide sites which lacked any potential oligosaccharide electron density, one, Asn117, is in the rarely glycosylated sequence Asn–Trp–Ser–Pro (Shakin-Eshleman *et al.*, 1996; Katsuri *et al.*, 1997; Gavel & von Heijne, 1990) and the other, Asn157 of $HEF_2$, is near the C-terminus of $HEF_2$ where the model *B* factors are high and the electron density is poor, suggesting that the disorder of this region is obscuring the oligosaccharide electron density.

Refinement was continued by including diffraction data to 3.2 Å resolution. A higher resolution native data set (an EMTS soak at 3 m*M*, Table 1, $\Delta F/F$ = 0.08) consisting of data between 5.5 and 3.2 Å were merged with the native data from 10 to 3.5 Å. The same set of reflections was used for the $R_{\rm free}$ set in the 10.0–3.5 Å range, but supplemented to include data from the 3.5–3.2 Å resolution shell. A cycle of positional refinement followed by simulated-annealing refinement with a bulk-solvent correction (*X-PLOR*; Brünger, 1992*b*) yielded an $R_{\rm free}$ of 26.7% and an $R_{\rm work}$ of 22.5% for the 10.0–3.2 Å resolution range (Table 4). In a final round of refinement, a maximum-likelihood target was used as implemented in *REFMAC* (Collaborative Computational Project, Number 4, 1994). $R_{\rm free}$ did not change, although $R_{\rm work}$ increased slightly.

The improvement in the electron density resulting from refinement is evident in Figs. 9(*d*) and 9(*h*). A section of the electron density from a simulated-annealing omit-map (Brünger, 1992*b*; Hodel *et al.*, 1992) with the model superimposed is shown in Fig. 12.

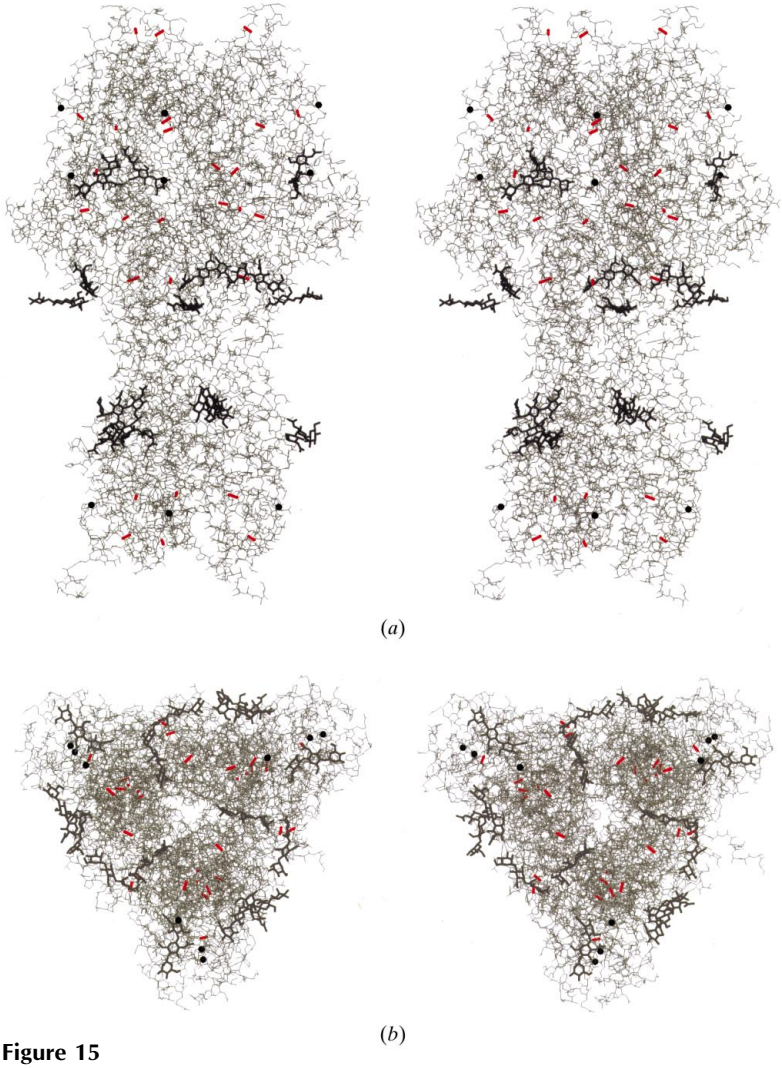



**Figure 15**
Stereographic drawing of the HEF trimer. (*a*) HEF trimer, all atoms, viewed perpendicular to the threefold-symmetry axis. The viral membrane is at the bottom. (*b*) HEF trimer viewed from the top of the molecule down the threefold-symmetry axis. Carbohydrate, bold. Carbohydrate sites without carbohydrate visible, dark dots. Disulfide bonds, red.

## 3. Discussion

### 3.1. Model assessment

The current model consists of residues 1–427 (of 432) for $HEF_1$ and residue 5–165 (of 175) for $HEF_2$. The final electron-density map has one major break in the main-chain electron density at residues 32–35 of $HEF_2$, where the sequence contains three consecutive glycines. There is poor electron density for 30 side chains, 26 of which are charged surface residues (labeled in Fig. 13). No model has been built for either the C-termini of $HEF_1$, residues 428–432, or of $HEF_2$, residues 166–175, owing to poor electron density in difference ($2F_o - F_c$) and omit electron-density maps. After iterated cycles of refinement and rebuilding, continuous but poor electron density was found in omit maps for residues 1–8 of $HEF_2$, which have very high $B$ factors (>100 Å$^2$). The correlation coefficient between the final electron-density map [($2F_o - F_c$), $\Phi_c$] and the model is 0.87 for the $HEF_1$ main chain and 0.86 for $HEF_1$ side chains (Fig. 13*a*). The correlation coefficient is somewhat lower for the $HEF_2$ chain: 0.82 for the main chain and 0.79 for the side chains (Fig. 13*b*). The Ramachandran plot (Fig. 14) shows that 76% of the residues have main-chain dihedral angles in the most-favored region and 0.2% are in the disallowed region.

### 3.2. HEF trimer structure

The HEF trimer has an extended stalk-like domain on the membrane proximal end and globular head-like domains at the membrane distal end (Fig. 15*a*). The three head domains are ∼60 Å tall and extend ∼116 Å in diameter at the widest point of the trimer (Fig. 15*b*). The head domains consist of only $HEF_1$ residues. They each contain a cell-receptor binding site and a cell-receptor-destroying enzyme active site (Fig. 16). The stem domain contains all of $HEF_2$ and part of $HEF_1$. It has three long $\alpha$-helices (60 Å) in the $HEF_2$ trimer interface and the membrane-fusion peptide at the N-terminal of $HEF_2$, located about 35 Å from the bottom of the molecule.

### 3.3. The $HEF_1$ polypeptide

The $HEF_1$ polypeptide chain starts from the bottom of the stem domain at the viral membrane end of the trimer and extends directly up to the head domain, except for a transverse $\beta$-hairpin loop (residue 14–28) halfway up the stem. $HEF_1$ reaches the top of the stalk at residue 43, 80 Å away from the bottom of the molecule (Fig. 17*a*). The polypeptide then forms part of the esterase domain (41–73) containing a $\beta$-strand followed by the active-site serine 57 (dot in Fig. 17*b*) and two short helical segments (Fig. 17*b*). A compact domain from residues 74–150 composes the top side of the esterase domain situated under the receptor-binding domain (Fig. 17*c*). One loop from this domain contributes one residue, Tyr127, to the receptor-binding site in the receptor-binding domain located directly above it. The receptor domain, residues 151–310, forms a 'jelly-roll' $\beta$-sandwich of nine $\beta$-strands at the distal end of the molecule (Fig. 17*d*). This domain contains the receptor-binding site (Fig. 16). Below the receptor-binding site, the polypeptide extends downward to complete the esterase domain. A compact, $\alpha\beta\beta\alpha$ motif (311–366), containing the active-site triad residues His355 and Asp352 (dots in Fig. 17*e*), packs against the N-terminal segments of the esterase, residues 41–73, which contained Ser57. The $HEF_1$ subunit ends in a segment, residues 367–432 (Fig. 17*f*), which runs down the stalk of the molecule, antiparallel to the N-terminal segment, residues 1–43. The five C-terminal

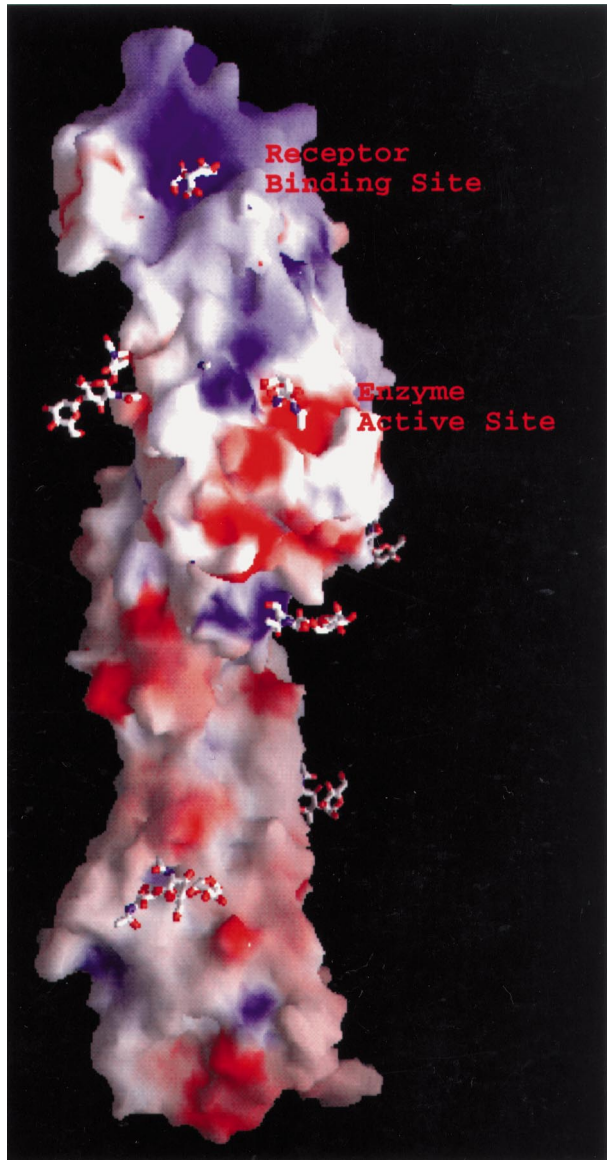



**Figure 16**
Surface diagram of HEF monomer. The receptor-binding pocket is located at the top of the head domain, while the enzyme active site is located near the widest part of the molecule. The electrostatic charge was calculated with *GRASP* (Nicholls *et al.*, 1991); red = negative, blue = positive. Also shown are the cell-receptor analogs bound to both the receptor-binding sites and the enzyme active sites. Trisaccharide models are shown at five of the eight potential glycosylation sites.

residues of $HEF_1$ are not observed in electron-density maps and are presumed to extend into the solvent and be disordered.

The topology of the $HEF_1$ subunit is dominated by $\beta$-strands (Fig. 18*a*). In the structure of $HEF_1$, there are six $\beta$-sheets arranged from bottom to top of the domain (colored in Fig. 18*b*). The two major segments of sequence difference in $HEF_1$ relative to $HA_1$, residues 41–73 and 311–366, which contain the catalytic triad of the esterase (Rosenthal *et al.*, 1998) are boxed in the topology diagram (Fig. 18*a*).

### 3.4. The $HEF_2$ polypeptide

$HEF_2$ is dominated by two antiparallel $\alpha$-helices, the first preceded by two antiparallel $\beta$-strands (Fig. 17*g*) and the second followed by two antiparallel $\beta$-strands (Fig. 17*h*). The first 24 residues which precede the two $\beta$-strands contain the non-polar fusion peptide implicated in the membrane-fusion activity of the HEF. The electron density for residues 1–5 of $HEF_2$ is poor probably owing to disorder of this region. Residues 13–18 of $HEF_2$ are sandwiched between N-terminal (9–13) and C-terminal (424–427) residues of $HEF_1$. The two antiparallel $\beta$-strands (residues 23–29 and 37–39) are part of a five-stranded $\beta$-sheet at the bottom of the HEF stem, the central strand of which is from the N-terminal of $HEF_1$ (residues 3–11; Figs. 17*g*, 17*h*; red in Fig. 18*b*). A $6\frac{2}{3}$-turn, 34 Å long $\alpha$-helix runs towards the top of the stem domain, followed by a 15-residue hairpin loop which reverses the direction of the chain (Fig. 17*g*). An 11-turn, ∼60 Å long $\alpha$-helix (residues 82–124) runs towards the viral membrane end of the stem. Near the end of the second $\beta$-strands following the long helix, Cys137 of $HEF_2$ is disulfide bonded to Cys6 of $HEF_1$, forming the only covalent link between $HEF_1$ and $HEF_2$ (Fig. 17*h*). An internal disulfide loop from $HEF_2$ 145–149 precedes a short $\alpha$-helix before the polypeptide turns toward the viral end of the molecule. The electron-density map could not be interpreted beyond residue 165. The transmembrane anchor starts about 20 residues after 165.

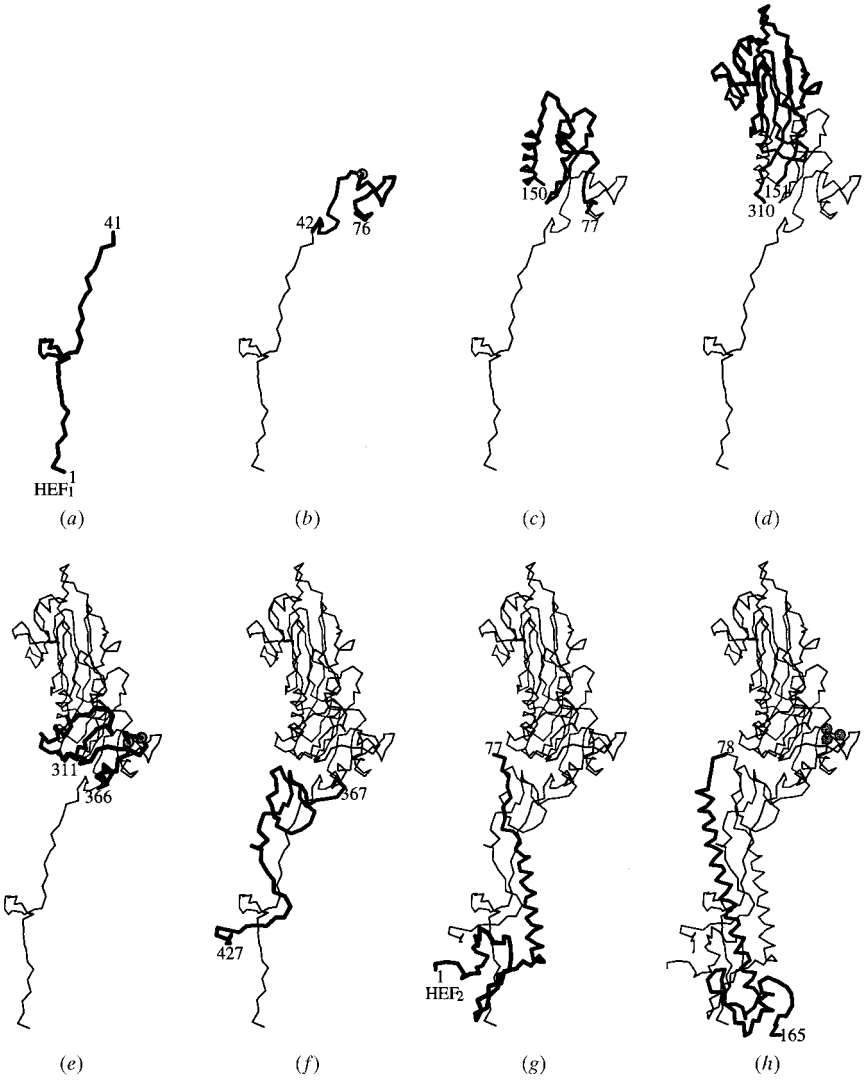



**Figure 17**
Polypeptide chain fold of HEF from N- to C-terminus. The catalytic triad, Ser57, Asp352 and His355, are shown as spheres in panels (*b*), (*e*) and (*h*).

### 3.5. The trimer interfaces

A single HEF monomer ($HEF_1$ + $HEF_2$) buries 4056 Å$^2$ of its solvent-accessible surface area in the HEF trimer interface. About 60% (2488 Å$^2$) of this buried surface is $HEF_2$ and 40% (1566 Å$^2$) is $HEF_1$. The prominent triple-stranded $\alpha$-helical interface at the core of the stalk domain accounts for only about 13% of the interface (321 Å$^2$) formed by $HEF_2$. This reflects the fact that the central helices diverge from each over both the top and bottom third of their lengths. The space on the threefold-symmetry axis between the top segments of the central helices is occupied by electron density which appears to be ions or water molecules (Rosenthal *et al.*, 1998). The space between the top of the two long $\alpha$-helices from adjacent $HEF_2$ monomers is filled by the long loop between the two helices of one $HEF_2$ monomer. This loop forms a substantial part of the $HEF_2$–$HEF_2$ interface (∼600 Å$^2$). The space between the long $\alpha$-helices of $HEF_2$ at bottom third of these helices is filled by the $HEF_2$ $\beta$-strand residues 133–135. After the low-pH-induced conformational change which activates the membrane-fusion activity, the loop and $\beta$-sheet residues forming these $HEF_2$–$HEF_2$ interfaces are expected to change positions radically, based on the precedent of the structurally related HA molecule (Bullough *et al.*, 1994).

The trimer interface formed by $HEF_1$ is also composed of three distinct regions. The

largest interface, 1240 Å$^2$, is at the top of the molecule, between the globular domains of $HEF_1$. The $HEF_1$ globular domain also contacts the $HEF_2$ subunit of the neighboring molecule, forming a 126 Å$^2$ interface with the extended loop between the two $\alpha$-helices of the $HEF_2$ subunit. A third contact (160 Å$^2$) is formed by the $HEF_1$ residues 21 and 22, which are part of a $\beta$-loop which projects toward the trimer axis in the stalk region of the molecule. These $HEF_1$ residues contact the $HEF_2$ subunit of an adjacent monomer at residue 106, which is part of the triple-stranded $\alpha$-helical coiled coil.

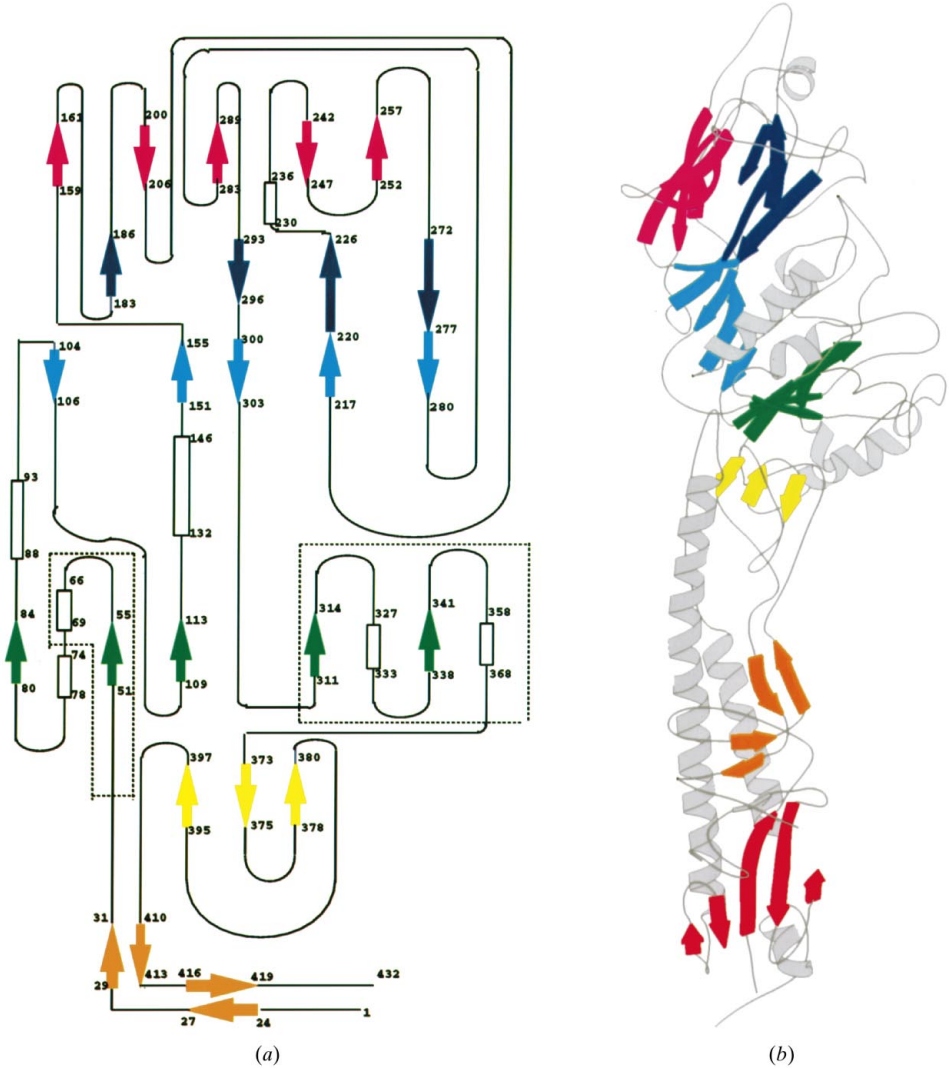
### 3.6. Oligosaccharides

The oligosaccharides seen at five ($HEF_1$ 12, 47, 130, 381 and $HEF_2$ 106) of the eight potential glycosylation sites (see above) appear to form three rings around the HEF trimer (Fig. 15*a*), one at the height of the fusion peptide, one near the junction of the top globular domain with the stem domain and one just below the receptor-binding site. The oligosaccharide attached at $HEF_2$ residue 157, which is in an area of poor electron density (see §2.8), occurs near the bottom of the trimer (Fig. 15*a*).

The oligosaccharide at $HEF_2$ residue 106 (on the long $\alpha$-helix) emerges from near the trimer axis and forms part of the trimer interface in the stalk of HEF (Figs. 10, 15*b*, 15*c*). The oligosaccharides linked at $HEF_1$ residues 47 and 130 also both extend toward other monomers, suggesting a possible role in stabilizing the trimer interface in the globular head region (Fig. 15*b*).

### 3.7. Separate receptor binding and esterase sites

The receptor-binding site is located near the top of the molecule, while the enzyme active site is located at the widest part of the trimer, near the base of the globular head domain (Fig. 16). Both sites are obvious cavities in the surface of the molecule (Fig. 16). Detailed structures of these sites and their interactions with a fragment of the cell receptor and an enzyme inhibitor are described elsewhere (Rosenthal *et al.*, 1998).




(*a*) (*b*)

**Figure 18**
$\beta$-Sheets in HEF. (*a*) Topology diagram of $HEF_1$, $\beta$-sheets are shown as colored arrows, $\alpha$-helices as cylinders. (*b*) HEF monomer showing the $\beta$-sheets and helices.

### 3.8. Domain structure and evolution

Three compact domains, labeled R, E, and F, can be identified in a Go distance plot (Go, 1981) of the HEF monomer (Fig. 19*a*). Only the R domain, which contains the receptor-binding site (Rosenthal *et al.*, 1998), is formed by a continuous segment of the HEF sequence. The extended F (membrane-fusion) domain which forms the stem of HEF is composed of three segments labeled F1, F2 and F3 in Fig. 19(*a*) and the esterase domain E is composed of E1, E′ and E2. Both the R and E domains (blue and green in Fig. 19*b*) have the topologies of large loops, with their respective N- and C-termini at almost the same point in the three-dimensional structure. The single-segment R domain is inserted into a surface loop of the esterase domain, and the esterase domain is inserted into a surface loop near the top of the stem domain, F (Rosenthal *et al.*, 1998).

A search of the Dali structure database with the HEF enzyme domain, E, identified structural similarity to the esterase from *Streptomyces scabies* (Wei *et al.*, 1995; *Z* score = 8.0, sequence identity 10%) and a brain acetylhydrolase (Ho *et al.*, 1997; *Z* score =

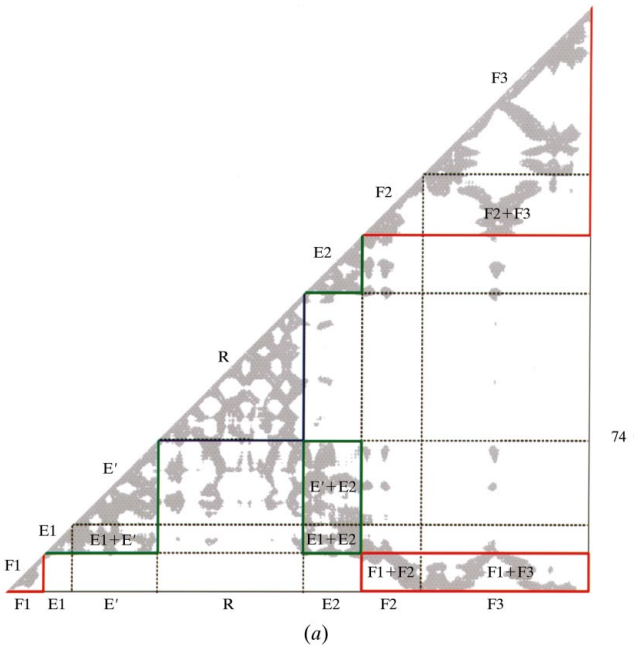



**Figure 19**
Domain structure of HEF. (*a*) Go distance plot (Go, 1981) of the HEF monomer. Three domains R, E, and F are outlined in blue, green and red, respectively. Filled areas are inter-C$\alpha$ distances of 22 Å or less. E and F are composed of non-contiguous sequences labeled F1, F2 and F3, and E1, E′ and E2. (*b*) HEF monomer colored by domain: blue = R, green = E, red = F. The stem domain F is elongated in the monomer, but compact as a trimer (see Fig. 6*c* in Rosenthal *et al.*, 1998).

8.3, sequence identity 13%; for details, see Rosenthal *et al.*, 1998). The implications of the domain structure of HEF for both the evolution of membrane-fusion proteins and for the evolution of influenza A, B, and C virus glycoproteins, as well as the HE (haemagglutinin-esterase) glycoprotein of coronaviruses has been described elsewhere (Rosenthal *et al.*, 1998).

We thank Mia Frayser and Richard Crouse for excellent technical assistance, one reviewer for suggesting additional molecular-replacement calculations and members of the Harrison–Wiley laboratory and the staff of the Cornell High Energy Synchrotron Source for assistance with data collection. XZ was supported by the Howard Hughes Medical Institute (HHMI). PBR was supported by Graduate Research Assistant Support from the Department of Molecular and Cellular Biology, Harvard University, as a graduate student and by a National Institutes of Health (NIH) training grant on the Molecular Basis of Viral Infectivity as a postdoctoral fellow. This research was supported by the NIH, The Deutsche Forschungsgemeinschaft, the Medical Research Council (UK) and the HHMI. DCW is an investigator of the HHMI.

## References

Air, G. M. & Compans, R. W. (1983). *Genetics of Influenza Viruses*, edited by P. Palese & D. W. Kingsbury, ch. 9. New York: Springer.
Braig, K., Adams, P. D. & Brünger, A. T. (1995). *Nature Struct. Biol.* **2**, 1083–1095.
Brünger, A. T. (1990). *Acta Cryst.* A**46**, 46–57.
Brünger, A. T. (1992*a*). *Nature (London)*, **355**, 472–475.
Brünger, A. T. (1992*b*). *X-PLOR Version 3.1 Manual*. New Haven: Yale University.
Bullough, P. A., Hughson, F. M., Skehel, J. J. & Wiley, D. C. (1994). *Nature (London)*, **371**, 37–43.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cornelissen, L. A. H. M., Wierda, C. M. H., van der Meer, F. J., Herrewegh, A. A. P. M., Horzinek, M. C., Egberink, H. F. & de Groot, R. J. (1997). *J. Virol.* **71**, 5277–5286.
Cowtan, K. D. (1994). *Jnt CCP4/ESF-EACBM Newslett. Protein Crystallogr.* **31**, 34–38.
Fitz, W., Rosenthal, P. B. & Wong, C.-H. (1996). *Bioorg. Med. Chem.* **4**, 1349–1353.
Gavel, Y. & von Heijne, G. (1990). *Protein Eng.* **3**, 433–442.
Go, M. (1981). *Nature (London)*, **291**, 90–92.
Hayes, B. K. & Varki, A. (1989). *J. Biol. Chem.* **264**, 19443–19448.
Herrler, G., Gross, H.-J., Imhof, A., Brossmer, R., Milles, G. & Paulson, J. C. (1992). *J. Biol. Chem.* **267**, 12501–12505.
Herrler, G., Multhaup, G., Beyreuther, K. & Klenk, H.-D. (1988). *Arch. Virol.* **102**, 269–274.
Herrler, G., Nagele, A., Heier-Ewert, H., Bhown, A. S. & Compans, R. W. (1981). *Virology*, **113**, 439–451.
Hewat, E. A., Cusack, S., Ruigrok, R. W. H. & Verwey, C. (1984). *J. Mol. Biol.* **175**, 175–193.
Ho, Y. S., Swenson, L., Derewenda, U., Serre, L., Wei, Y., Dauter, Z., Hattori, M., Adachi, T., Aoki, J., Arai, H., Inoue, K. & Derewenda, Z. S. (1997). *Nature (London)*, **385**, 89–93.
Hodel, A., Kim, S.-H. & Brünger, A. T. (1992). *Acta Cryst.* A**48**, 851–858.
Imhof, A., Herrler, G. & Brossmer, R. (1988). *Biol. Chem. Hoppe-Seyler*, **369**, 841–842.
Jones, E. Y., Walker, N. P. C. & Stuart, D. I. (1991). *Acta Cryst.* A**47**, 753–770.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Katagiri, S., Ohizumi, A. & Homma, M. (1983). *J. Infect. Dis.* **148**, 51–56.
Katsuri, L., Chen, H. & Shakin-Eshleman, S. H. (1997). *Biochemistry*, **323**, 415–419.
King, B., Potts, B. J. & Brian, D. A. (1985). *Virus Res.* **2**, 53–59.
Kleywegt, G. J. & Jones, T. A. (1994). *Proceedings of the CCP4 Study Weekend. From First Map to Final Model*, edited by S. Bailey, R. Hubbard & D. A. Waller, pp. 59–66. Warrington: Daresbury Laboratory.
Kleywegt, G. J. & Jones, T. A. (1996*a*). *Structure*, **4**, 897–904.
Kleywegt, G. J. & Jones, T. A. (1996*b*). *Acta Cryst.* D**52**, 826–828.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Luytjes, W., Brendenbeek, P. J., Noten, A. F. H., Horzinek, M. C. & Spaan, W. J. M. (1988). *Virology*, **50**, 118–124.
McIntosh, K., Kapitikian, A. A., Turner, H. C., Hartley, J. W. & Chanock, R. M. (1970). *Am. J. Epidemiol.* **91**, 585–592.
Madden, D. R. (1992). PhD thesis, Harvard University, USA.
Monto, A. S. & Lim, S. K. (1974). *J. Infect. Dis.* **129**, 271–276.
Nakada, S., Creager, R. S., Krystal, M., Aaronson, R. P. & Palese, P. (1984). *J. Virol.* **50**, 118–124.
Navaza, J. (1994). *Acta Cryst.* A**50,** 157–163.

Nicholls, A., Sharp, K. A. & Honig, B. (1991). *Proteins*, **11**, 281–296.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pfeifer, J. B. & Compans, R. (1984). *Virus Res.* **1**, 281–296.
Pleschka, S., Klenk, H.-D. & Herrler, G. (1995). *J. Gen. Virol.* **76**, 2529–2537.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Rosenthal, P. B. (1996). PhD thesis, Harvard University, USA.
Rosenthal, P. B., Formanowski, F., Treharne, A. C., Newman, J., Skehel, J. J., Meier-Ewert, H. & Wiley, D. C. (1996). *Acta Cryst.* D**52**, 1041–1045.
Rosenthal, P. B., Zhang, X., Formanowski, F., Fitz, W., Wong, C.-H., Meier-Ewert, H., Skehel, J. J. & Wiley, D. C. (1998). *Nature (London)*, **396**, 92–96.
Shakin-Eshleman, S. H., Spitalnik, S. L. & Kasturi, L. (1996). *J. Biol. Chem.* **271**, 6363–6366.
Vlasak, R., Luyjes, W., Leider, J., Spaain, W. & Palese, P. (1988). *J. Virol.* **62**, 4686–4690.
Vlasak, R., Muster, T., Lauro, A. M., Powers, J. C. & Palese, P. (1989). *J. Virol.* **63**, 2056–2062.
Wei, Y., Schottel, J. L., Derewenda, U., Swenson, L., Patkar, S. & Derewenda, Z. S. (1995). *Nature Struct. Biol.* **2**, 218–223.
Weis, W. I., Brünger, A. T., Skehel, J. J. & Wiley, D. C. (1990). *J. Mol. Biol.* **212**, 737–761.
Wiley, D. C., Wilson, I. A. & Skehel, J. J. (1981). *Nature (London)*, **289**, 373–378.
Wilson, I. A., Skehel, J. J. & Wiley, D. C. (1981). *Nature (London)*, **289**, 366–373.