Supporting Information

Money et al. 10.1073/pnas.0805740106

SI Materials and Methods

CD Analysis. CD data (190–250 nm) was collected by using a Jasco J-810 Spectropolarimeter, and analyzed by using the CDPro package available via Dichroweb service (http://www.cryst.bbk.ac.uk/cdweb/html/home.html). Each of the 3 programs within CDPro (Contin/LL, CDSSTR, and Selcon3) uses a different algorithm for deconvoluting experimental data with reference to a training dataset consisting of proteins of known structure for which CD data are available. All 3 programs were tested to determine the most suitable for processing the M^{254R} data. Given that M^{254R} is an extrinsic membrane protein, the selection of the most appropriate training dataset was considered. Two datasets were investigated: SP43 (Fig. S3) consists only of soluble proteins; SMP56 (Fig. S4) contains examples of membrane proteins. From the data presented below, it is apparent that Contin/LL produces the best fit to the experimental data, and there is little difference between the results produced by the 2 datasets. The Contin/LL/SP43 data are reported in the main text.

Cloning of pETM^{254R}. The full-length protein ORF of the respiratory syncytial virus matrix protein (RSV M) gene was cloned by PCR as previously described (1). Briefly, RNA was isolated from RSV (A2 strain) infected VERO cells by using an Ambion RNA extraction kit, and used to prepare cDNA by using random hexameric primers. The M ORF was amplified in a PCR by using gene specific primers (sequence available on request) that contained XhoI restriction enzyme sites designed to facilitate cloning into the pET16a vector from Novagen. The cloning resulted in the formation of an in-frame fusion with the histidine tag and a Factor Xa proteolytic cleavage sequence derived from the vector. Subsequently, it was determined that a base misincorporation in the synthesis of the primers representing the 3′ end of the M ORF or a mutation occurred during cloning resulted in a methionine to arginine substitution. This clone was designated pETM^{254R}.

Expression of M^{254R} Protein. The pETM^{254R} plasmid was used to transform *Escherichia coli*, strain BL21 CodonPlus (Novagen), for expression purposes. Protein was expressed using an autoinduction system (for a full description and recipes of media used, see ref. 2). Briefly, an isolated colony of BL21 (CodonPlus) containing pETM^{254R} was grown overnight in a noninducing medium (MDG; a number of websites define media suitable for autoinduction, based on Studier formulation; the recipe for MDG can be found at http://strgen.org/protocols/studier.pdf as September 2007) at 37 °C and used to inoculate an autoinduction medium (ZY 5052). The diluted culture (1:1,000) was grown for 8 h at 37 °C before reducing the temperature to 22 °C for a further 24 h at 22 °C, and the OD_{600} monitored every hour after 18 h. The culture was harvested when saturation was achieved (OD_{600} of $\approx$ 10) by centrifugation and the pellet used directly for protein extraction.

Expression of Selenomethionine Labeled M^{254R} Protein. A 1:1,000 dilution of BL21 (CodonPlus) containing pETM^{254R} from an overnight MDG culture used to inoculate the defined autoinduction medium PASM 5052 that contained selenomethionine. The original recipe (2) called for a methionine:selenomethionine ratio of 1:100; however, it was found empirically that a 1:200 ratio was more effective. The M^{254R} protein was expressed as in ZY 5052 medium, with the exception that saturation and induction of the culture did not occur until after 96 h at 22 °C. Cells were harvested as described above and processed for protein extraction.

Isolation of M^{254R} Protein. Cell pellets containing either methionine or selenomethionine labeled protein were suspended in phosphate buffer (50 mM NaH_2PO_4/300 mM NaCl, pH 7.4) and lysed by sonication after pretreatment with lysozyme in the presence of RNase A (100 $\mu g \cdot mL^{-1}$) and DNase I (10 $\mu g \cdot mL^{-1}$). After sonication, lysates were made 8 mM with respect to the detergent CHAPS (Sigma), and cell debris was pelleted at 50,000 $\times g$ for 30 min at 4 °C. The clarified supernatant was made 50 mM with respect to imadazole and applied to a nickel affinity column (Qiagen). The column was washed with increasing amounts of imadazole in phosphate buffer, and the protein was found to elute once the gradient was approximately 150 mM. Fractions containing M^{254R} protein were pooled and dialyzed into the appropriate buffer for downstream processing. To remove the histidine-tag, proteins were exposed to the proteolytic enzyme Factor Xa in cleavage buffer (0.005 unit per μg protein/100 mM NaCl/50 mM Tris·HCl/5 mM CaCl2, pH 8.0) for 48 h at 21 °C. Factor Xa was removed by using the Factor Xa Cleavage Capture Kit (Novagen) as per manufacturer's directions. The released histidine-tag, and any noncleaved protein were removed by Ni affinity chromatography. Each stage of the preparation of M^{254R} was monitored by SDS/PAGE.

Crystallization and X-Ray Data Collection. Crystallization conditions for M^{254R} were screened by using the sitting drop vapor diffusion method, together with the Hampton Crystal screen, Crystal screen 2, Hampton Index screen (Hampton research), and the PACT premier screen (Molecular Dimensions); 100 nL droplets were formed by the Mosquito (TTP LabTech Ltd) liquid handling robot, mixing 50 nL protein with 50 nL of well solution. Initial crystals were obtained in condition 29 of the Index screen 70% (vol/vol) Tacsimate pH7. These conditions were optimized further to improve crystal quality resulting in conditions of between 55 and 65% (vol/vol) Tacsimate pH7. A cryoprotectant solution was produced by supplementing the mother liquor with an additional 30% (vol/vol) glycerol. The flattened needle shaped crystals were harvested in rayon fibre loops then bathed in cryoprotectant solution before flash freezing in liquid N_2. Seleno-methionine substituted crystals were grown in the same manner. The mercury derivative was obtained by soaking a seleno-methionine substituted crystal for $\approx$15 min in 50% (vol/vol) Tacsimate solution containing 10 mM thiomersal (ethylmercury thiosalicylate) buffered to pH 7 with 0.1 M Tris. The crystals were then back soaked in cryoprotectant before flash freezing in liquid N_2. Data were collected on single crystals of native protein at the European Synchrotron Radiation Facility (Grenoble, France) at 100 K with a $\Delta\varphi$ of 0.5°, over an oscillation range of 360° in inverse collection mode on ID14.4 by using a ADSC Q4R CCD at a wavelength of 0.97950 Å. Data for the seleno-methionine derivative of M^{254R} were collected at Stanford Synchrotron Radiation Laboratory (SSRL) on station 9.2 with a $\Delta\varphi$ of 1° over an oscillation range of 260° at a wavelength of 0.97922 Å by using a MarMosaic-325 CCD detector. The appropriate wavelength for data collection at the Se edge, optimized for the f′′ component of the anomalous scattering, was determined by using a fluorescence scan. Mercury derivative data were also collected by using station 9.2 at the SSRL with a $\Delta\varphi$ of 1° and over an oscillation range of 180° at a wavelength of 1.00473 Å. The appropriate energy for data collection at the Hg

L2 edge optimized for the f′′ component of the anomalous scattering was determined by using a fluorescence scan. Low energy remote data for the Se-derivatized protein at 120 K were collected by using an in-house copper rotating anode source (Rigaku Micromax 007HF) at a wavelength of 1.54179 Å over an oscillation range of 145° with a $\Delta\varphi$ of 0.5° by using a Mar345 image plate detector. All data were indexed and integrated in MOSFLM (3); all other computing was carried out in the CCP4 suite unless otherwise stated (4). All crystals of M were found to crystallize in the monoclinic space group C2 with unit cell parameters of a = 52.30 Å, b = 78.70 Å, c = 66.02 Å, and β = 96.28° with 1 molecule in the asymmetric unit. The structure was solved by using MIRAS on all 4 derivatives in autoSHARP (5); 5% of the native data were set aside for cross validation analysis, and the behavior of R_{free} was used to monitor and guide the refinement protocols. ARP/wARP (6) in conjunction with REFMAC was used to automatically build the sequence into the electron density (7). Refinement was undertaken in REFMAC with manual correction to the model by using COOT (8). Coordinates and observed structure factors have been deposited with the Protein Data Bank (accession codes 2vqp for the coordinates and r2vqpsf for the structure factors). X-ray data and refinement quality statistics are shown in Table S1.

Circular Dichroism Spectroscopy. Protein samples were dialysed against 5 mM phosphate buffer overnight at 4 °C. Far-UV CD spectra and the corresponding blanks were recorded between 190 and 250 nm in a cuvette of path length 0.2 cm by using a Jasco J-810 Spectropolarimeter by averaging 8 accumulations recorded at a rate of 10 nm/min, with a pitch of 0.5 nm, a bandwidth of 1 nm, and a response time of 2 s. Near-UV spectra were recorded by using a 1-cm cell, with a pitch 0.2 nm, and a response time of 1 s. After subtraction of the appropriate blank, binominal smoothing was carried out within the Jasco Spectra Analysis program. Smoothed data were analyzed for protein secondary structure by using the CDSSTR, SELCON3, and CONTIN/LL programs (9), accessed either via the Dichroweb service (10, 11) or the CDPro package. Both a general protein (SP43/dataset 4) and a membrane protein (SMP56/dataset 10) reference set were used (9, 12). Data collected included wavelengths down to 190 nm, which significantly increases in the comparison of CD derived structural information for comparison with crystallographic data (13).

Protein Cross-Linking. In a typical experiment, a solution of M^{254R} protein (14 μM) and glutaraldehyde (0.41 M) in 0.6 mL of water was incubated for 2 min at room temperature before the addition of sodium borohydride (1 M in 0.1 M NaOH; 90 μL). The solution was then dialysed overnight against water and analyzed by MALDI-TOF mass spectrometry by using a sinapinic acid matrix on a Bruker Daltonics Autoflex TOF/TOF instrument.

Sequence Analysis of Related Proteins. The sequences of 8 pneumoviruses were selected for analysis: gi:3089376 (human RSV, A2 strain), gi:60549168 (human RSV, Long strain), gi:2582027 (RSV, B1 strain), gi:9631272 (bovine RSV), gi:3123017 (ovine RSV), gi:49823137 (Avian Metapneumovirus), gi:46310178 (Pneumovirus of Mice), and gi:34420899 (Human Metapneumovirus). Analyzes were performed on all 8 protein sequences and a subset containing only the RSV proteins. Amino acid sequences were aligned to the RSV M sequence with the program ClustalX (1.83.1) (14) by using the Gonnet weight matrix and secondary structure-based penalties obtained from the RSV M structure. Other parameters were as default. Column quality scores were analyzed, with cut-offs of 20 for the full sequence set and 20 or 50 for the RSV-only sequence set used to identify nonconserved residues.

1. Henderson G, Murray J, Yeo RP (2002) Sorting of the respiratory syncytial virus matrix protein into detergent-resistant structures is dependent on cell-surface expression of the glycoproteins. *Virology* 300:244–254.
2. Studier FW (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expression Purif* 41:207–234.
3. Leslie AGW (1992) *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography* (Daresbury Laboratory, Warrington, UK), Vol 26.
4. Collaborative Computational Project Number 4 (1994) *Acta Crystallogr D* 50:760–763.
5. Vonrhein C, Blanc E, Roversi P, Bricogne G (2006) Automated Structure Solution With autoSHARP. *Methods Mol Biol* 364:215–230.
6. Perrakis A, Morris R, Lamzin VS (1999) Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* 6:458–463.
7. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* 53:240–255.
8. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D* 60:2126–2132.
9. Sreerama N, Woody RW (2000) Estimation of protein secondary structure from circular dichroism spectra: Comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set *Anal Biochem* 287:252–260.
10. Lobley A, Whitmore L, Wallace BA (2002) DICHROWEB: An interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics* 18:211–212.
11. Whitmore L, Wallace BA (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res* 32:W668–W673.
12. Sreerama N, Woody RW (2004) On the analysis of membrane protein circular dichroism spectra. *Protein Sci* 13:100–112.
13. Greenfield NJ (2006) Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols* 1:2876–2890.
14. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882.

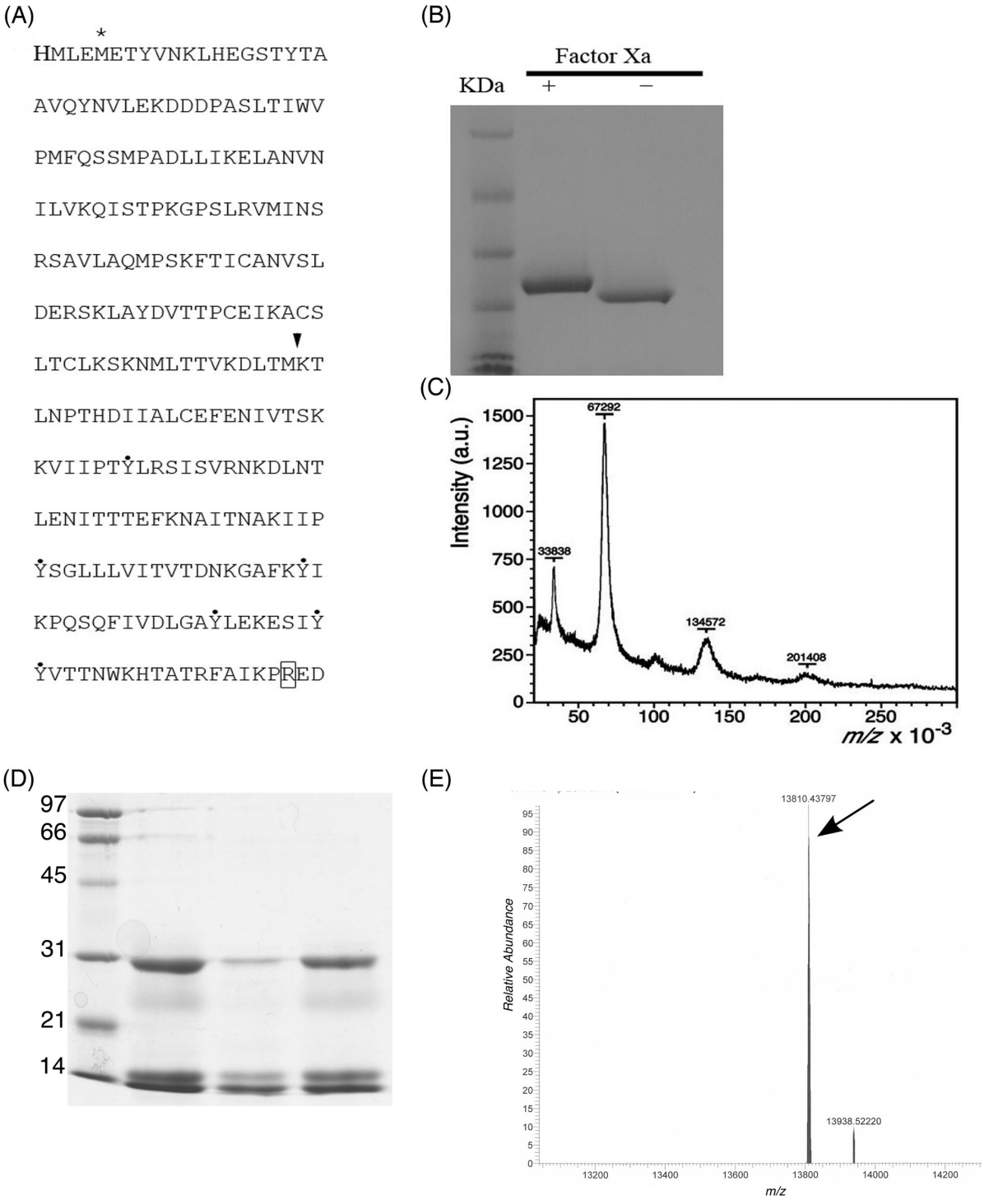


Fig. S1. Purification and cross-linking experiment on M^{254R}. (*A*) Sequence of the protein successfully crystallized. The protein has 4 residues (HMLE) at the amino terminus that are derived from the expression vector (pET16b, Novagen), but which are not resolved in the crystal structure. The initiating methionine of M is indicated by an asterisk, and all numbering in main text is taken from this residue as 1. The site of cleavage (see main text), as mapped by mass spectrometry, is indicated by a black arrowhead, and the residue 254R is boxed. The tyrosine residues in the C-terminal domain that are presented in a planar ring are indicated by black dots. (*B*) Purified M^{254R} protein before (−) and after (+) Factor Xa cleavage. (*C*) Mass spectrum of the material isolated after cross-linking of RSV M (14 μM) with glutaraldehyde (410 mM) in PBS (pH 7.4). The presence of higher order forms of the protein can be observed; particularly relevant is the occurrence of hexamers. (*D*) Gel showing 2 protein fragments at ≈14 kDa that form upon prolonged storage of M254R. Molecular weight markers were run in the left lane and are annotated with their corresponding mass in kDa. (*E*) The major ion observed in MS analysis of the mixture corresponding to the gel shown in *D* is indicated by an arrow. This peak, with an observed *m/z* of 13810.44, correlates with a calculated *m/z* of 13810.80 for cleavage between K and T (see *A*) and is also seen in samples with N-terminal His tags.

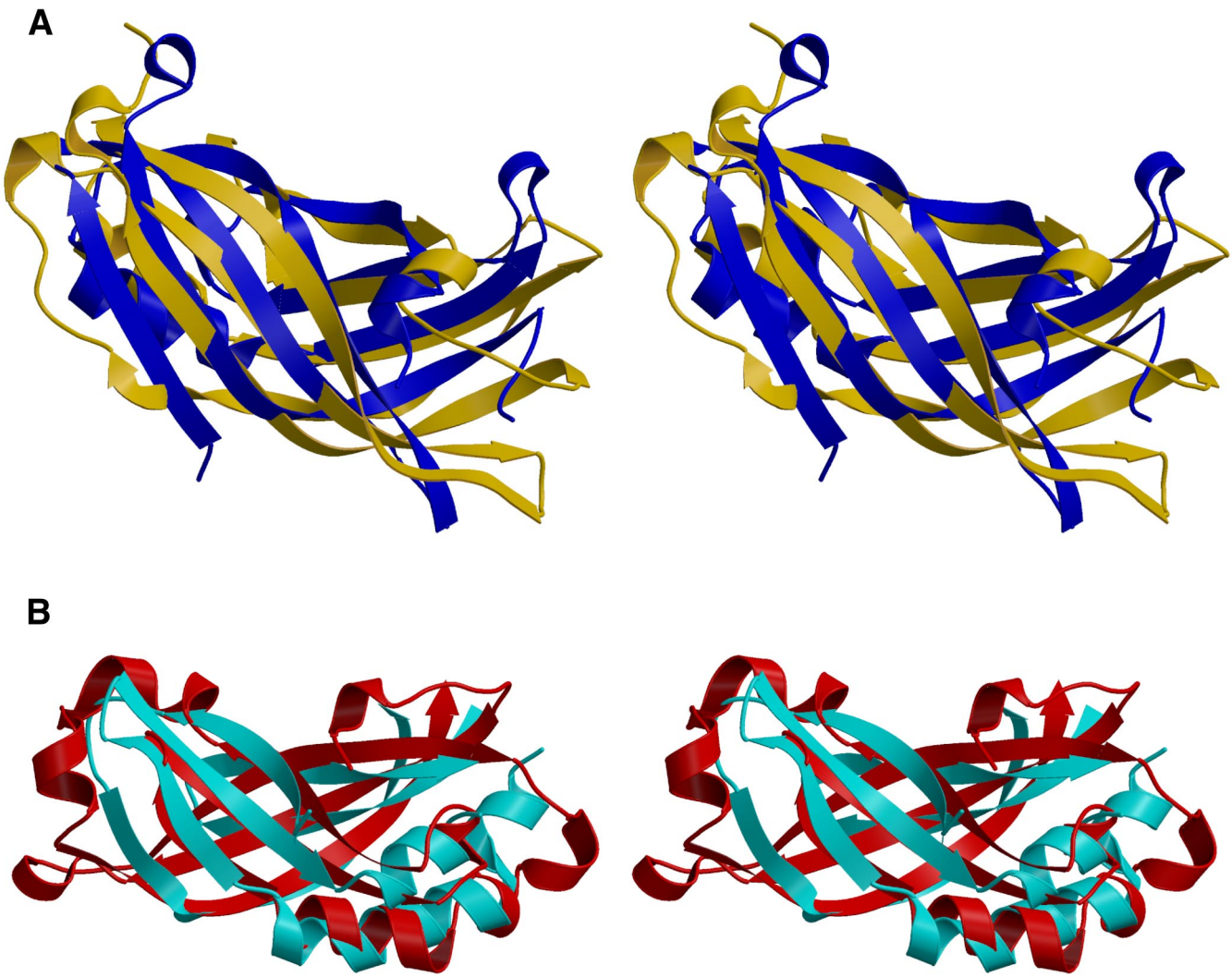


Fig. S2. Stereoscopic views of the superimposition of M254R and EBOV VP40 N- and C-terminal domains. Domains from M254R and EBOV VP40 (PDB code 1ES6) were modeled and compared separately to account for differences in relative domain position. (*A*) RSV M protein, M^{254R}, N-terminal domain in blue and VP40 in yellow. (*B*) M^{254R} protein C-terminal domain in red and VP40 in cyan.

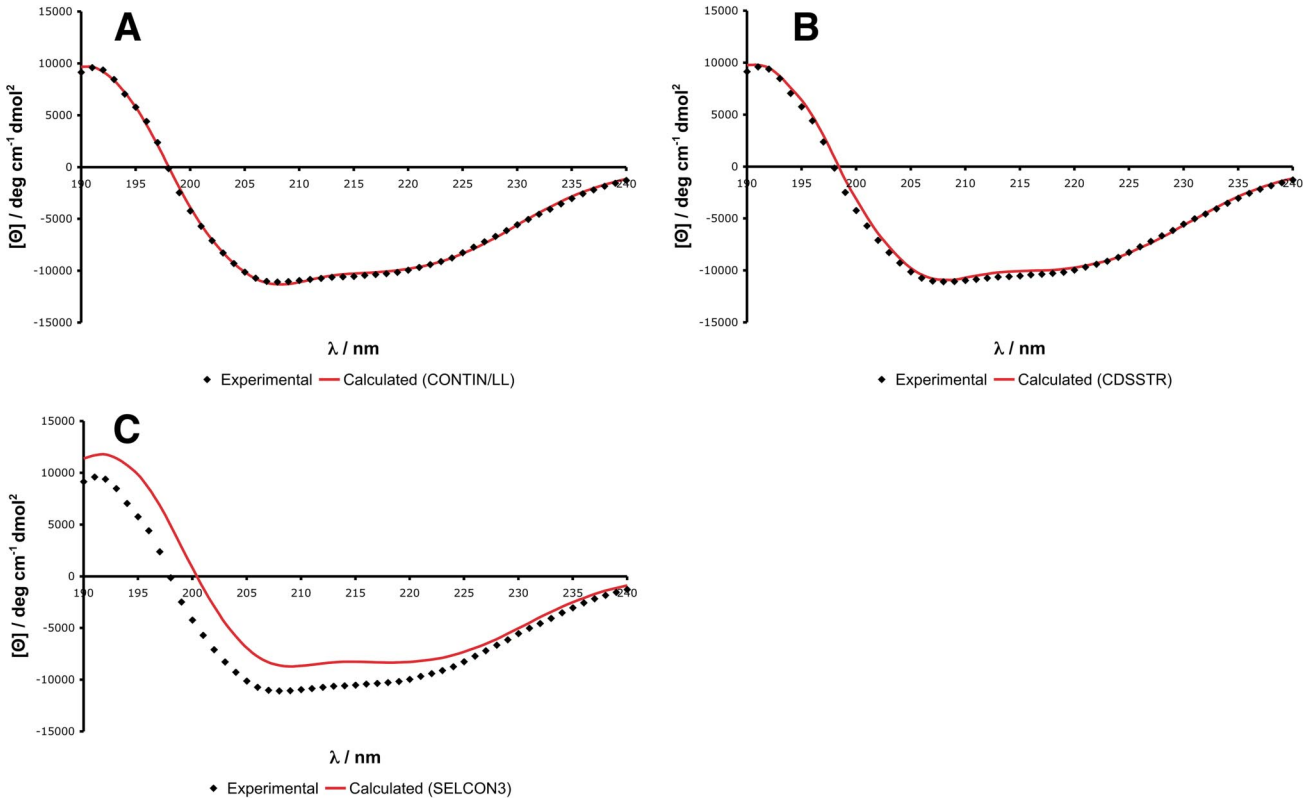


Fig. S3. Experimental and calculated CD spectra for hRSV M^{254R} (I). CD spectrum (190–250 nm) was obtained from M^{254R} in 5 mM phosphate buffer (pH 7.4) at 20 °C. Calculated spectra were obtained by using the CONTIN/LL (*A*), CDSSTR (*B*), and Selcon3 (*C*) algorithms with the SP43 protein dataset as reference.

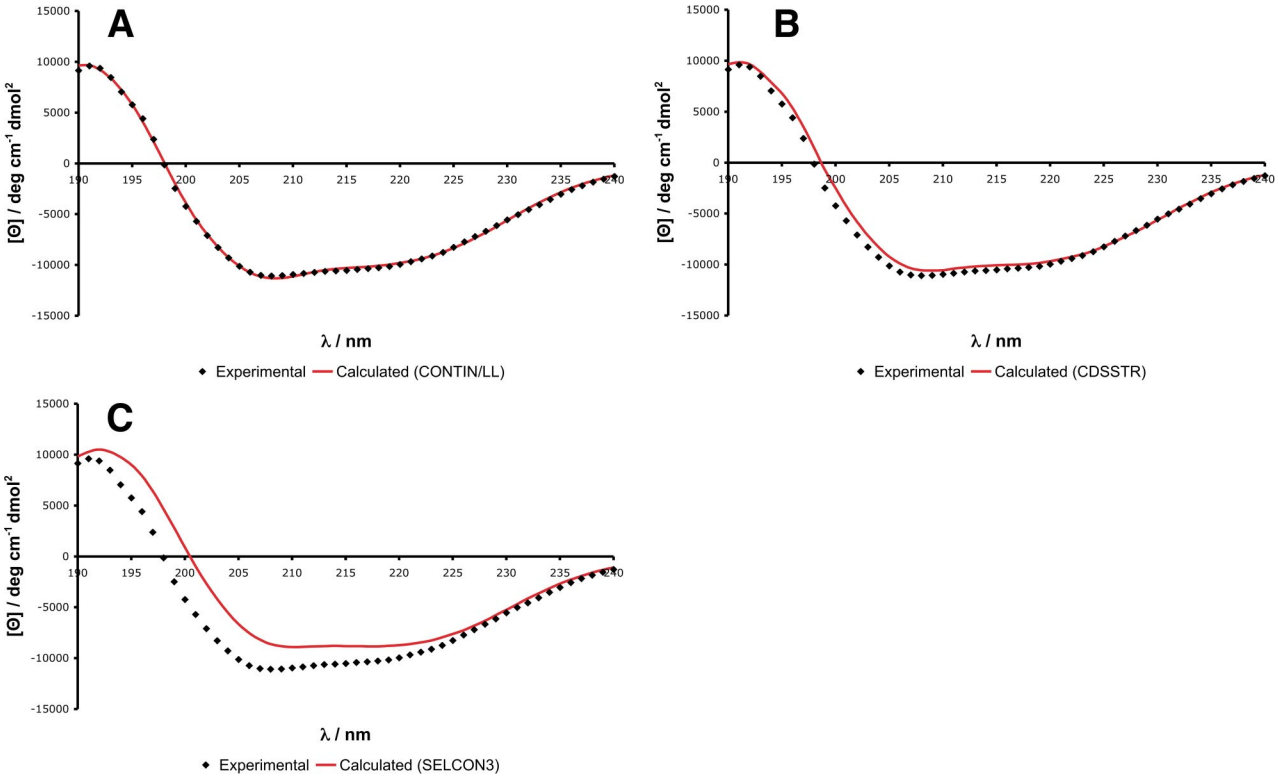


Fig. S4. Experimental and calculated CD spectra for hRSV-M protein (II). CD spectrum (190–250 nm) was obtained from M^{254R} in 5 mM phosphate buffer (pH 7.4) at 20 °C. Calculated spectra were obtained using the CONTIN/LL (*A*), CDSSTR (*B*), and Selcon3 (*C*) algorithms with the SMP56 protein dataset as reference.

(A)

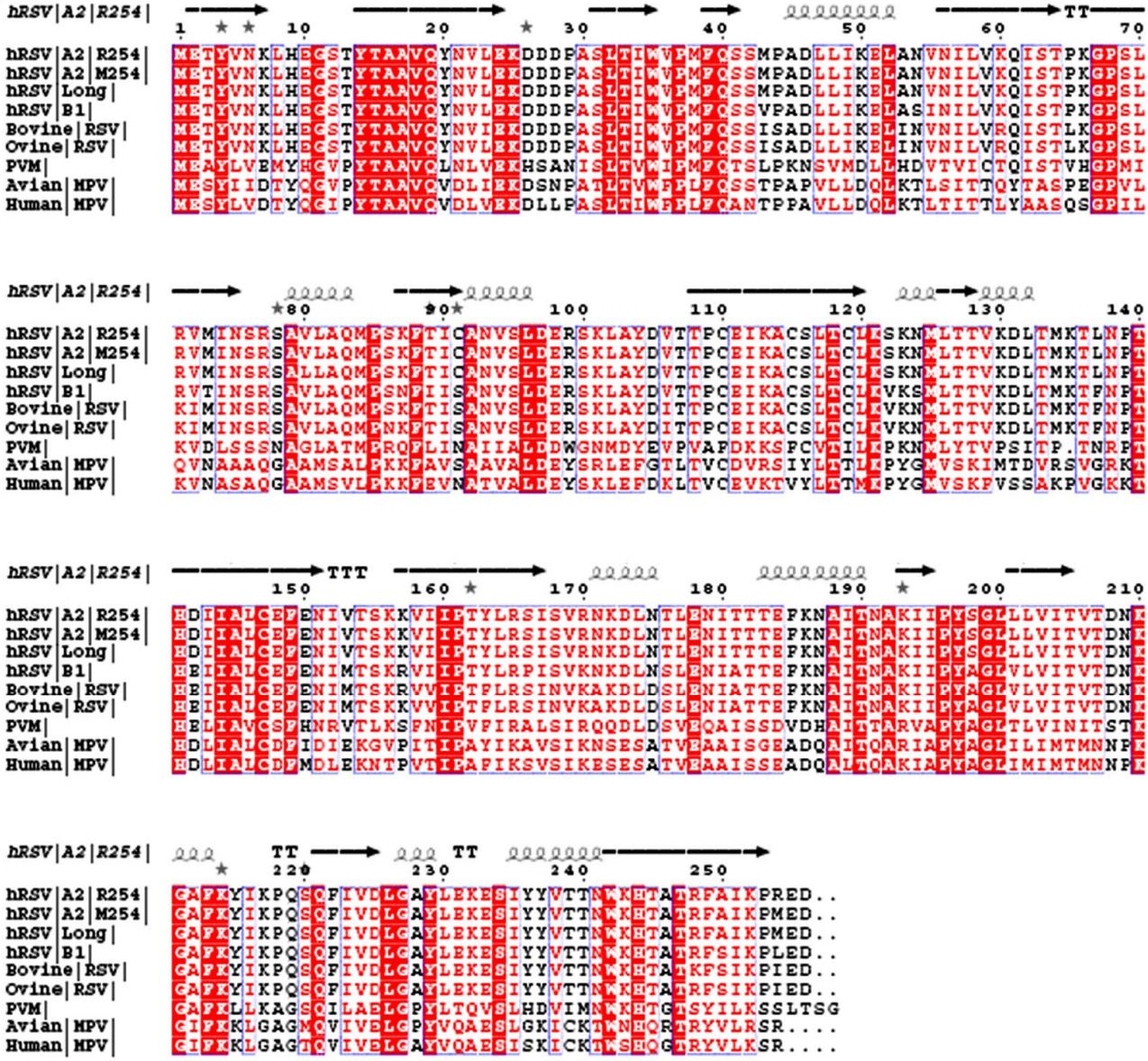
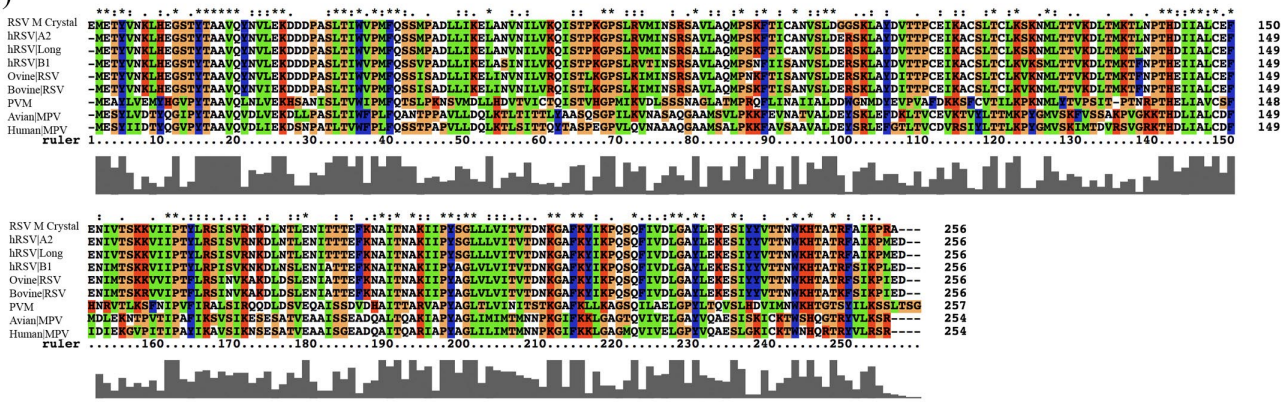


Fig. S5. Computed structure/sequence comparison of pneumoviral Ms. (*A*) Alignment of representative pneumovirus sequences with the derived structure of RSV A2 M. The sequences were aligned by using initially ClustalW and the output file, with the M^{254R} pdb file were submitted to the ESPrit server (http://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi) for alignment based on the structure. Sequences boxed in red (white text) are conserved, residues in red (white background) are conservative in nature, those in black show strain specific variation. Arrows represent β-sheets and coils, α-helices. The high degree of conservation and the subdivision of the pneumoviruses (i.e., RSV) and the metapneumoviruses (PVM, hMPV, and aMPV) are readily apparent. (*B*) Alignment of pneumoviral M with a Q score of 20 for the full sequence set; and (*C*) 20 or 50 for the RSV-only sequence set used to identify nonconserved residues. To perform the alignments, a Gonnet weight matrix was used, and secondary structure-based penalties were obtained from the RSV M structure. The information from such alignments was used to identify surface residues (Fig. 4) that may be important for species specific interactions, i.e., with the nucleocapsids and F proteins. Such information may predict which residues would be important for future mutagenesis experiments.

(B)



(C)

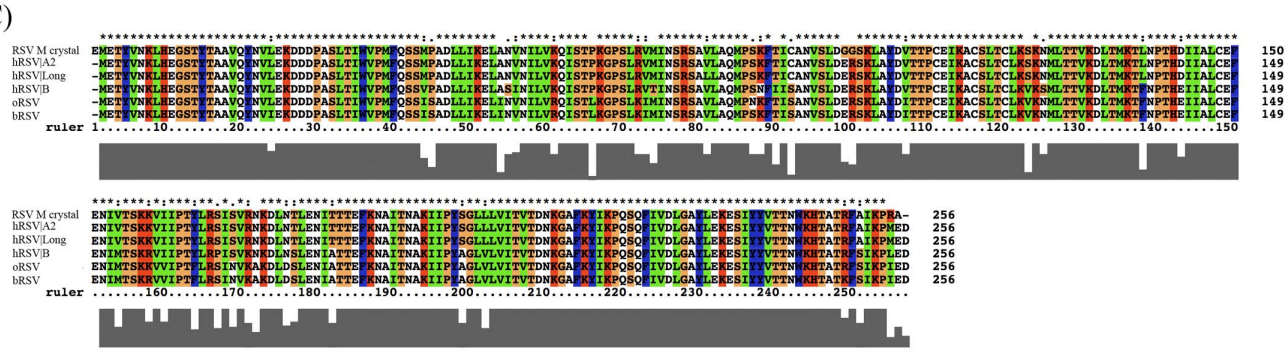


Fig. S5 (continued).

Table S1. X-ray data collection and refinement statistics

	Native M collected at ESRF	SeMet low energy remote	SeMet peak	Hg soak peak
Beamline	ID14.4	N/A	BL9.2	BL9.2
Wavelength, Å	0.97950	1.54179	0.97922	1.00473
Resolution of data	65.65–1.60 (1.68–1.60)	27.359–2.500 (2.64–2.50)	65.65–2.00 (2.11–2.00)	65.938–2.700 (2.85–2.7)
R_{merge}	0.062 (0.506)	0.081 (0.251)	0.093 (0.393)	0.184 (0.55)
Mean I/σI	23.4 (3.5)	18.1 (5.6)	14.0 (4.6)	8.2 (2.0)
Completeness, %	96.2 (96.2)	96.4 (96.4)	100 (100)	98.4 (98.4)
Multiplicity	8.2 (8.3)	5.1 (5.1)	5.5 (5.4)	3.7 (3.8)
R_{cryst}	0.190	—	—	—
R_{free}	0.23	—	—	—
rmsd 1–2 bonds, Å	0.17	—	—	—
rmsd 1–3 bonds, °	1.693	—	—	—
Avg main chain B, Å^2	20.54	—	—	—
Avg side chain B, Å^2	23.23	—	—	—
Avg substrate B, Å^2	—	—	—	—
Avg solvent B, Å^2	37.66	—	—	—
PDB code	2vqp	—	—	—

Of modelled residues, 88.9% are found in the ″most favoured″ regions of the Ramachandran plot, 11.1% in the ″allowed″ region, and no residues lie in the generously allowed or disallowed areas. In addition to the amino acid chain, a number of heteromolecules are also visible in the electron density map; these are derived either from the crystallization conditions or from the solution used for cryo-protection. The additional molecules consist of 2 glycerol molecules, 7 formate ions, and 3 acetate ions. N/A, not applicable. Numbers in parentheses refer to the highest resolution shell.

Table S2. Output of X-ray structural analysis

	α-Helix	3_{10}-Helix	Strand	Turn	Coil	Bridge
No. residues	30	21	120	53	30	2
% of total	11.7	8.2	46.9	20.7	11.7	0.8

Summary of the data from analysis of the M^{254R} structure using the program Stride.

Table S3. CDPro output data

	Structure	H(r)	H(d)	S(r)	S(d)	Turn	Unrd	rmsd (Exp-Calc)	nrmsd (Exp-Cal)
Fractions	SELCON3	0.129	0.119	0.136	0.097	0.222	0.309	0.761	0.315
	CONTIN/LL	0.119	0.121	0.131	0.097	0.224	0.309	0.045	0.019
	CDSSTR	0.125	0.122	0.135	0.090	0.216	0.319	0.234	0.097

Three programs (Contin/LL, CDSSTR, and Selcon3) were used with the SP43 protein dataset. Data are extracted from the file ProtSS.out.

Table S4. CDPro output data

	Structure	H(r)	H(d)	S(r)	S(d)	Turn	Unrd	rmsd (Exp-Calc)	nrmsd (Exp-Cal)
Fractions	SELCON3	0.124	0.117	0.139	0.095	0.222	0.320	0.710	0.294
	CONTIN/LL	0.123	0.117	0.128	0.092	0.223	0.317	0.056	0.023
	CDSSTR	0.119	0.113	0.138	0.090	0.217	0.321	0.205	0.085

Three programs (Contin/LL, CDSSTR, and Selcon3) were used with the SMP56 protein dataset. Data are extracted from the file ProtSS.out.