Supporting Online Material

Co-Folding Organizes the AMV RNA and Coat Protein for Replication

Laura M. Guogas, David J. Filman, James M. Hogle and Lee Gehrke

Materials and Methods

Sample Preparation and Crystallization

The 39 nucleotide minimal RNA (Fig. S1A) was synthesized chemically to include 5-Bromo-U at positions U858 and U872 (Dharmacon Research). The RNA was purified by denaturing polyacrylamide gel electrophoresis (PAGE), recovered by electroelution in an Elutrap chamber (Schleicher and Schuell), concentrated by ethanol precipitation, and de-salted over a Sep-Pack column (Waters). Peptide CP26 (Fig. S1B) was chemically synthesized (MIT Biopolymers) and purified by reversed phase HPLC over a C18 column (Vydac) by methanol gradient.

Crystals of the AMV RNA peptide complex were grown by hanging drop vapor diffusion at 22°C over a reservoir containing 2.45 M ammonium sulfate, 10mM magnesium acetate, and 50mM MES buffer pH 5.6. An RNA-peptide binding solution containing 0.5 μl of RNA (4 mg/ml) was mixed with a twofold molar ratio of coat protein peptide. A 2:1 stoichiometry was suggested by biochemical data (*1*). The RNA and peptide reagents were sequentially added to a solution containing 10mM Tris pH 7.5, 1mM DTT, 50 mM NaCl, and 1mM $SmCl_3$ and allowed to stand for 5 minutes at RT to form the RNA-peptide complex. One μl of this solution was then mixed in equal volume

with reservoir solution and then suspended over the reservoir. The crystals grew as parallelopipeds (~100 x 100 x 40 μm^3) in 5-7 days. Prior to data collection, crystals were cryoprotected by stepwise addition of a harvest solution containing 2.6 M ammonium sulfate, 10mM magnesium acetate, 1mM $SmCl_3$, 50mM MES buffer pH 5.6, and 15% glycerol. The crystals were then flash frozen in liquid nitrogen.

X-ray data collection, phasing and structure refinement

MAD data collection was performed at the Structural Biology Center, line ID-19 of the Advanced Photon Source (Argonne, IL) and processed with DENZO and SCALEPACK from the HKL2000 package (*2*). The crystals are in space group $C222_1$ with unit cell dimensions a = 50.1 Å, b = 123.0 Å, c = 53.6 Å. Phase information was obtained from two crystals. For the first crystal, diffraction data were collected at the peak, inflection point and a low energy remote setting near the K-edge of bromine. A highly redundant peak data set was collected from the second crystal (Table S1). Data were collected to 2.9 Å; however, the quality diminished precipitously beyond 3 Å, causing us to base the refinement on the 3 Å dataset. The two Br atoms were located in the asymmetric unit using the program SOLVE (*3*); however the occupancy of the second site, at U858, was low. Refinement of the Br positions and phase calculations to 3.0 Å were performed using the program SHARP (*4*). Solvent flipping was done using the program SOLOMON (*5*). The resultant maps were suitable to begin model building. Positioning of an A-form RNA helix was initially done using the positions of the Br atoms to indicate the register of the sequence. The AMV atomic model was manually built into the experimental density using the program XtalView/Xfit (*6*). Model building

was alternated with refinement of the model using the maximum likelihood method in REFMAC 5.1 (*7*). The experimental phases were maintained as strong restraints throughout the structure solution. The model was gradually improved by manual inspection and judged an improvement by a decrease in R_{free} during refinement. Translation, libration and screw-rotation (TLS) refinement was performed using REFMAC 5.1 with the entire complex considered as a single pseudo-rigid body. Late in refinement, X-PLOR (*8*) was used briefly to enforce the co-planarity and hydrogen bonding of specific base pairs before resuming TLS refinement in REFMAC 5.1. Very late in refinement, the overfitting of atomic parameters to errors in the data was corrected by introducing 0.1 to 0.4 Å random noise (using PDBSET in CCP4) into the atomic coordinates prior to maximum likelihood refinement with REFMAC 5.1 The final refinement produced an R_{cryst} of 24.9 and R_{free} of 26.8 (Table S1). We were unable to place all of the 5' nucleotides (853-855) from the first stem loop and the extreme N-terminus of the CP peptides into the model. Information for the 3' stem loop nucleotides is also not available due to the formation of an RNA dimer.

Electrophoretic Mobility Shift Assay (EMSA)

Band shift assays were performed as previously described (*9, 10*). For testing dimer formation, $AMV_{843\text{-}881}$ (39mer) and $AMV_{815\text{-}881}$ (67mer) RNA oligonucleotides were synthesized by *in vitro* transcription, renatured, and 20 nM RNA was mixed with 500nM full length coat protein and loaded onto a 10% polyacrylamide gel. For testing AUGC mutations, 39mer RNA oligonucleotides were synthesized by *in vitro*

transcription, renatured, and 20 nM RNA was mixed with 500 nM coat protein peptide CP26 and loaded onto a 12% polyacrylamide gel.

Supplementary Text

The crystallographic dimer

The $AMV_{843-881}$ RNA in complex with CP26 coat protein peptide was crystallized in space group $C222_1$ with one monomer per asymmetric unit (or one half dimer, Fig. S1 A). The unusually high concentration of RNA present in crystallization experiments promoted the formation of a dimeric complex containing two molecules of $AMV_{843-881}$RNA and four CP26 peptides (Fig. S2 A and B). Each of the two $AMV_{843-881}$ RNA monomers forms an authentic 5' hairpin structure. However, instead of forming the 3' stem loop, two symmetry-related 3' tails form Watson-Crick pairs with one another.

Prior biochemical data did not predict the formation of an RNA dimer (*10*), but the observation of the dimer in the crystal structure prompted us to perform an additional electrophoretic mobility gel shift assay (EMSA) (*9*) to assess the potential for RNA dimer formation in solution (Fig. S2C). 39mer and 67mer RNAs comprising the minimal binding site $AMV_{843-881}$ or a longer fragment, $AMV_{815-881,}$ were mixed with coat protein either alone or in combination. Lanes 1 and 2 show the 39mer RNA alone and with a 10-fold molar excess of coat protein, while lanes 3 and 4 show the 67mer alone and with coat protein. In each case, the data suggest that a single species of free RNA was present (lanes 1 and 3) while one (lane 2) or multiple (lane 4) species of RNA-coat protein complexes were observed. We anticipated that the RNAs were not forming dimers in

solution due to the presence of only one band in the RNA only samples (lanes 1 and 3). However, it is possible that these single bands represent homodimers of each of the RNA species. Therefore, in lanes 5 and 6 we mixed the two RNA species. If RNA dimers were forming we expected to see a heterodimeric RNA species in lane 5, or possibly if the presence of coat protein is needed, a heterodimeric RNA-coat protein complex in lane 6. We did not observe these complexes in either case. Lane 5 shows 2 bands with the same mobilities of each of the RNAs in lane 1 and 3, while lane six also contains bands seen previously in lanes 2 and 4. Prior biochemical experiments also support the presence of a monomeric RNA-protein species. Ansel-McKinney used chemical footprinting to examine the conformation of bound and unbound $AMV_{843\text{-}881}$ RNA. In both cases, nucleotides predicted to lie in loop positions (AMV 854-867, the 5' loop and AMV 872-874, the 3' loop) were susceptible to digestion by the single strand specific ribonuclease T2, indicating that these are single stranded regions (*11*). If a dimer formed at the 3' end, then nucleotides 872 and 874 would not be susceptible to T2 digestion because they would be base paired. Additionally, changes in the putative 3' stem loop sequence that would make RNA-RNA dimerization less favorable because loop nucleotides would not be able to base pair, did not affect coat protein binding (*1*). Note that the base pairs in the 3' stem of the dimer are identical to those expected in the 3' stem of the monomer. In this sense each half of the dimer can be thought of as a monomeric complex with a large insertion in the loop of the 3' hairpin (Fig. S2B). We therefore argue that it is reasonable to discuss the structure in terms of the monomeric unit. Further justification for this argument is provided by the striking correlation of the structure and biochemical data (*1*, *12*, *13*). In the crystal the two molecules of the dimer

are related by a crystallographic twofold axis (arrow, Fig. S2A). To extract the monomer, we have divided the molecule such that the monomer consists of the 5’ hairpin and the first half of the 3’ helix (residues 843-871) from one RNA and second half of the 3’ hairpin and terminal AUGC (residues 875-881) from the second RNA (Fig. S2B).

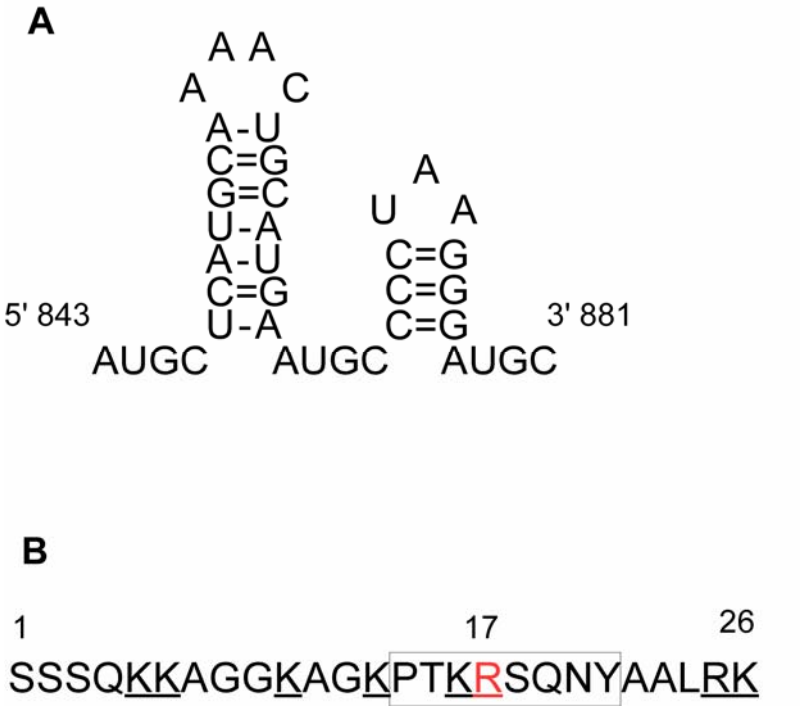
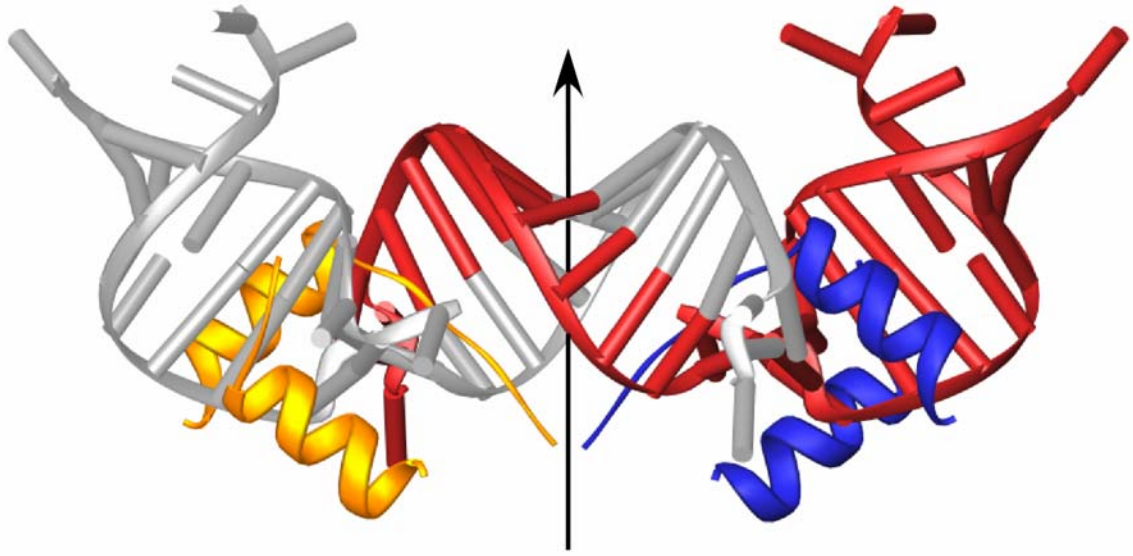
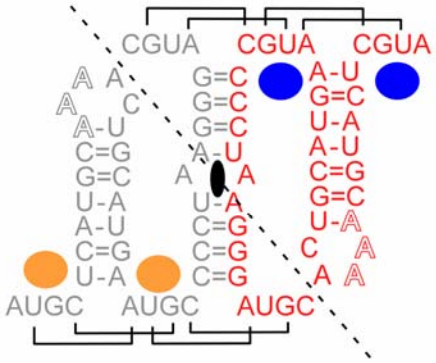


Figure S1. The AMV RNA and coat protein peptide. (A) The 39-nucleotide minimal coat protein binding domain found at the 3' ends of AMV RNAs, $AMV_{843\text{-}881}$. (B) The N-terminal 26 amino acids of the viral coat protein, CP26. Basic residues are underlined, R17 is highlighted in red, and boxed residues represent the PTxRSxxY RNA binding consensus sequence.

A



B


CGUA
CGUA
CGUA
AUGC
AUGC
AUGC

C

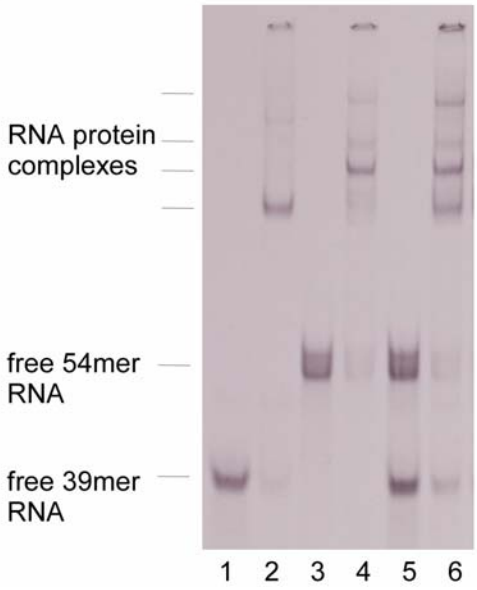

RNA protein complexes
free 54mer RNA
free 39mer RNA
1
2
3
4
5
6

Figure S2: The crystallographic dimer. (A) Ribbon diagram of the crystallographic dimer. One RNA monomer is shown in grey and the second in red. Peptides associated with each RNA are shown in gold and blue respectively. The arrow shows the twofold symmetry axis. (B) Diagram showing the association of the two RNA strands. Each RNA helix is extended by two base pairs from the intervening AUGC sequences, indicated by brackets. Nucleotides shown in outline were not ordered in the model. The small black oval represents the twofold symmetry axis, while the dashed line divides the dimer into two monomeric units. (C) 39mer and 67mer RNAs comprising the minimal binding region, AMV843-881, and the minimal binding region plus an additional 5' hairpin, AMV815-881 were used in an EMSA assay. Lane 1: Free 39mer RNA, Lane 2: 39mer RNA plus coat protein; Lane 3: free 67mer RNA; Lane 4: 67mer RNA plus coat protein; Lane 5: 39mer and 67mer RNA; Lane 6: 39mer and 67mer RNA plus coat protein. 20 nM RNA (all lanes) was incubated with 500 nM coat protein (even lanes).

A

```
AMV 3/4      GCUAACGCACAUAUAUAAAUGCUCAUGCAAAACUGCAUGAAUGCCCCUAAGGGAUGC
CiLRV 3/4    GUGCGUAGAUGCCUAUAUUUUCUCUCCUGAGAAAAUAUAGAUGCCUCCAAAGGAGAUGC
CVV 3/4      AAUAUAGAUGCCCAAACUCUCUCUCAUGGAGAGAGAAUGGAUGCCUCCGAAGGAGAUGC
EMV 3/4      UAAUAGAUGCCUAAUUCUCUCUCUCAGGGAGAGAGAUUAGAUGCCUCCAAGGAGAUGC
PDV 3/4      UGCCUACAAAUUUUGUACAUGCCCUCACCGUAAGGUGAGGAUGCCCCUUUAAGGGAUGC
SLV 3/4      AUGUAGAUGCCUAAUACUCUCUCUCAGGGAGAGAGUUUAGAUGCCUCCAAAGGAGAUGC
TSV 3/4      AUGCUUACGUUUGGUGCCAGUAGUAUAUAAUAUACUACUGAUGCCUCCUUUAUAGGAGAUGC
PNRSV 3/4    CUUACCUGCGUUAGCAGAUGCCCACAACGUGAAGUUGUGGAUGCCCGUUAGGGAAGC
TAMV 3/4     UAGCUAAGUUCAUAUGCCCACCUUUGCUGUCUCCGGGUGGAUGCCUCAUGGUGCUAUGGAUGC
HdMV 3/4     UAAUAGAUGCCUAAUUCUCUCUCUCAGGGAGAGAGAUUAGAUGCCUCCAAGGAGAUGC
```

B

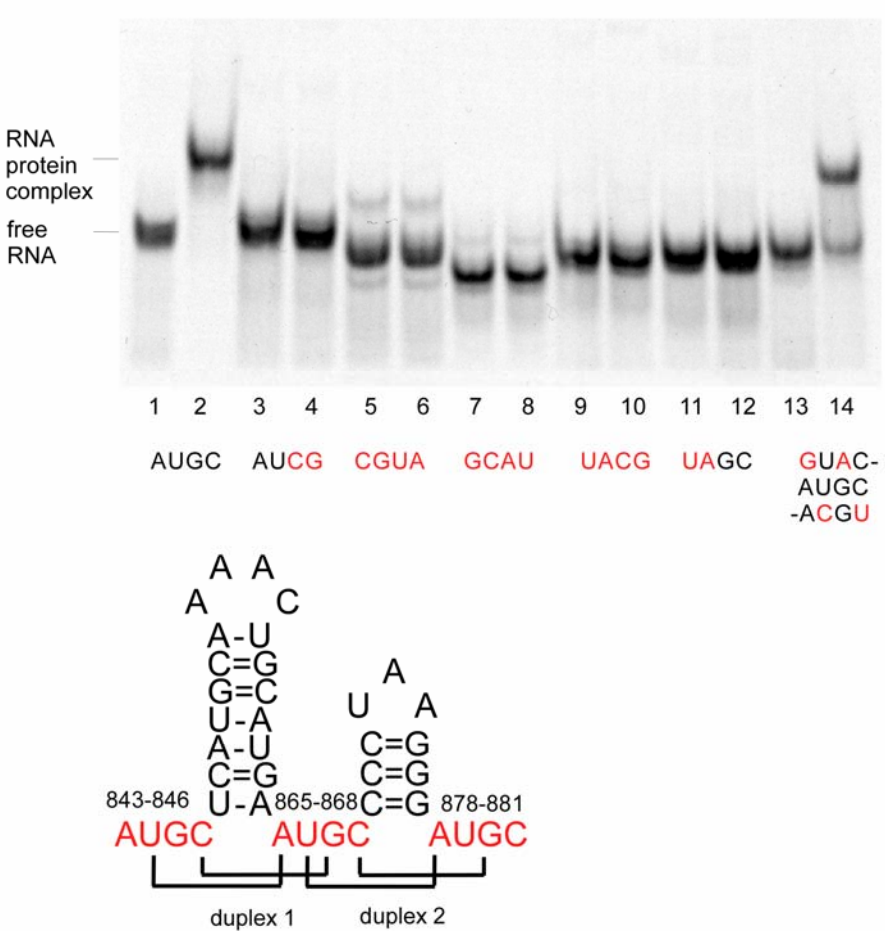


Figure S3: The AUGC repeats are required for CP binding. (A) Sequence alignment of AMV and ilarvirus 3' RNAs. Sequences were aligned against AMV865-868. Approximately 50 nucleotides comprising the extreme 3' end are shown for each virus, alfalfa mosaic virus (AMV), citrus leaf rugose virus (CiLRV), citrus variegation virus (CVV), elm mottle virus (EMV), prune dwarf virus (PDV), spinach leaf latent virus (SLV), tobacco streak virus (TSV), prunus necrotic ringspot virus (PNRSV), tulare apple mosaic virus (TAMV), hydrangea mosaic virus (HdMV). AUGC sequences are indicated in bold and underline. Terminal AAGC sequences are indicated in bold only. (B) EMSA bandshift with mutant 39mer RNAs. Lanes 1 and 2: wild type AMV 843-881 RNA. Lanes 3-12: Mutations, indicated in red, were made to each of the 3 AUGC sequences. Lanes 13-14: Mutations were made to the 5' and 3' AUCG sequences as shown in red. 20 nM RNA (all lanes) was incubated with 500 nM CP26 (even lanes). A schematic of AMV 843-881 RNA with the AUGC repeats that were mutated in red is shown at the bottom of the figure.

Table S1: Data collection, phasing and refinement statistics

Data Collection[a]

	Peak1	Infl1	Rem1	Peak2
Wavelength (Å)	0.91942	0.91997	0.92181	0.91942
Resolution (Å)	99-3.0 (3.19-3.0)	99-3.0 (3.19-3.0)	99-3.0 (3.19-3.0)	99-3.0 (3.19-3.0)
Total Measurements	47,235	41,427	44,613	69,733
Unique Reflections	3,509	3,487	3,495	3,559
Completeness (%)	99.8 (99.8)	99.4 (98.1)	99.7 (99.1)	99.6 (100.0)
Redundancy	13.5	11.9	12.8	19.6
Rsym(%)[b]	8.4 (41.2)	8.5 (40.5)	8.5 (44.5)	7.1 (25.1)
$\langle I \rangle / \langle \sigma(I) \rangle$	28.9 (5.7)	25.0 (4.4)	27.1 (5.1)	39.9 (11.8)

Phasing

Resolution (Å)	27-3.0	27-3.0	27-3.0	27-3.0
Isomorphous				
Rcullis (a/c)[c]	0.687/0.780	-	0.809/0.918	0.871/0.905
Phasing Power (a/c)[d]	0.986/0.828	-	0.443/0.374	0.265/0.236
Anomalous				
Rcullis (a)	0.863	0.968	0.996	0.776
Phasing Power (a)	0.867	0.363	0.083	1.211

Refinement

Resolution	20-3.0 (3.16-3.0)
Unique Reflections	3236 (453)
R_{cryst}[e]	0.249 (.285)
R_{free}[e]	0.268 (.267)
Average B factor (Å^2)	45.2
R.m.s.d. Bonds (Å)	0.013
R.m.s.d. Angles (°)	2.118

[a]Values in parentheses refer to data in the highest resolution shell.

[b]$R_{sym} = \Sigma_{hkl}|I_{hkl} - \langle I_{hkl} \rangle| / \Sigma_{hkl} \langle I_{hkl} \rangle$ where $\langle I_{hkl} \rangle$ is the mean intensity of symmetry related observations of a unique reflection

[c]Rcullis = <phase-integrated lack of closure> / < | Fph - Fp | >
(a/c) indicates acentric/centric. (a) indicates centric only.

[d]Ppower = < [| Fh(calc) | / phase-integrated lack of closure] >

[e]$R_{cryst} = \Sigma_{hkl}|F_{obs} - F_{calc}| / \Sigma_{hkl}|F_{obs}|$. $R_{free} = R_{cryst}$ for 9.2% of the data that were not used in refinement.

Supporting References and Notes

1. P. Ansel-McKinney, L. Gehrke, *J. Mol. Biol.* **278**, 767 (1998).

2. Z. Otwinowski, W. Minor, *Methods Enzymol.* **276**, 307 (1997).

3. T. C. Terwilliger, J. Berendzen, *Acta Crystallogr.* **D55**, 849 (1999).

4. E. de La Fortelle, G. Bricogne, *Methods Enzymol.* **276**, 307 (1997).

5. J. P. Abrahams, A. G. W. Leslie, *Acta Crystallogr.* **D52**, 30 (1996).

6. D. E. McRee, *J. Struct. Biol.* **125**, 156 (1999).

7. G. N. Murshudov, A. A. Vagin, E. J. Dodson, *Acta Crystallogr.* **D53**, 240 (1997).

8. A. T. Brunger, *X-PLOR, version 3.1. A system for X-ray crystallography and NMR* (Yale University Press, New Haven, CT, 1992).

9. F. Houser-Scott, M. L. Baer, K. F. Liem, Jr., J.-M. Cai, L. Gehrke, *J. Virol.* **68**, 2194 (1994).

10. M. Baer, F. Houser, L. S. Loesch-Fries, L. Gehrke, *EMBO J.* **13**, 727 (1994).

11. P. Ansel-McKinney, Ph.D., Harvard University (1996).

12. P. Ansel-McKinney, S. W. Scott, M. Swanson, X. Ge, L. Gehrke, *EMBO J.* **15**, 5077 (1996).

13. G. Rocheleau, J. Petrillo, L. Guogas, L. Gehrke, *J. Virol.* **78**, 8036 (2004).

Supporting Online Material
www.sciencemag.org
Materials and Methods
Supplementary Text
Figs. S1, S2, S3
Table S1