# Large-Scale Sequence Analysis of Avian Influenza Isolates

John C. Obenauer,[1] Jackie Denson,[1] Perdeep K. Mehta,[1] Xiaoping Su,[1] Suraj Mukatira,[1] David B. Finkelstein,[1] Xiequn Xu,[1] Jinhua Wang,[1] Jing Ma,[1] Yiping Fan,[1] Karen M. Rakestraw,[1] Robert G. Webster,[2,4] Erich Hoffmann,[2] Scott Krauss,[2] Jie Zheng,[3] Ziwei Zhang,[3] Clayton W. Naeve[1,4]*

The spread of H5N1 avian influenza viruses (AIVs) from China to Europe has raised global concern about their potential to infect humans and cause a pandemic. In spite of their substantial threat to human health, remarkably little AIV whole-genome information is available. We report here a preliminary analysis of the first large-scale sequencing of AIVs, including 2196 AIV genes and 169 complete genomes. We combine this new information with public AIV data to identify new gene alleles, persistent genotypes, compensatory mutations, and a potential virulence determinant.

Influenza A viruses are endemic in the wild aquatic birds of the world and are occasionally transmitted to humans with catastrophic results (1). The outbreak of H5N1 avian influenza virus (AIV) infection in humans in Southeast Asia starting in 2003 has resulted in 76 deaths among 144 infected individuals (2) and the slaughter of millions of birds. The virus is clearly present in the wild duck population in China (3, 4) and has spread to Romania, Turkey (5), Croatia, and Russia. These events have led to substantial global concern about the potential for this virus to evolve to pandemic proportions, with the capacity to cause millions of deaths. Remarkably, little AIV whole-genome data are available, and the public repositories of sequence data are skewed toward the shorter genes (Matrix and Nonstructural) and surface glycoproteins [hemagglutinin (HA) and neuraminidase (NA)]. A more comprehensive collection of data and analysis at the gene and whole-genome level are critical needs as we search for answers to questions regarding the virulence and transmissibility of these viruses from avian species to humans.

To address this need, we established the first large-scale sequencing effort to collect additional genomic data on the avian population of influenza A viruses. The AIV genome consists of eight RNA segments that range in size from 890 to 2341 bases and code for 11 known proteins. We developed a primer library specific for each of these eight genes and suitable for our remarkably diverse virus collection, as well as a pipeline for assembly, finishing, and quality control (6). Our sequences are of high quality, averaging 8.8 reads per base and mean

[1]Hartwell Center for Bioinformatics and Biotechnology, [2]Department of Infectious Diseases, Division of Virology, [3]Department of Structural Biology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA. [4]Department of Pathology, University of Tennessee Health Science Center, Memphis, TN 38163, USA.

*To whom correspondence should be addressed. E-mail: clayton.naeve@stjude.org

Phred quality values of 51.2, indicating an average base-calling accuracy of 99.999% (7). We provide complete gene sequences, including 5' and 3' untranslated regions. All sequences produced by the St. Jude Influenza Genome Project have been submitted to GenBank and are available for study.

**Study population and sequencing results.** The St. Jude Influenza Repository currently contains ~11,000 influenza viruses, including ~7000 AIVs. We have sequenced a diverse sampling of 336 AIVs from this collection, including isolates from ducks, gulls, shorebirds, and poultry collected in North American, Eurasian, and Australasian countries, primarily during the years from 1976 to 2004. Our sampling includes representatives of all 25 known HA and NA serotypes. We report here the analysis of 2196 new AIV gene sequences and 169 complete AIV genomes (listed in table S1), doubling

the amount of publicly available genetic information for AIVs (3,702,178 bases of finished data). We also include in our analysis 2143 AIV sequences retrieved from GenBank (table S2) for a total of 4339 AIV genes. We calculated (Table 1) the nonsynonymous/synonymous substitution rate ratios (dN/dS) in our population (8) and find that only the gene encoding the recently reported alternative PB1 transcript PB1-F2 (9) is under positive selection pressure. The ratio for PB1-F2 is extreme; values greater than 1.0 are considered positive selection, and this gene's ratio is 9.36; furthermore, this open reading frame is conserved in 281 of 284 of the PB1 sequences in our study. Thus, this protein's reported role in apoptosis appears to be critical. We calculated the number of variants per position in the amino acid sequence of each protein and find that avian viruses exhibit greater variability in their PB1-F2, HA, NS1, and NA proteins than do their non-avian counterparts (table S2 lists the accession numbers for the avian and non-avian samples). An analysis of concatenated genomes showed that HA, NA, and NS1 contribute the most to the variability of avian virus genomes, and that multiple viral lineages co-circulate (supporting online material text). HA and NA are considered highly variable because of immune pressure; it is not known what drives NS1 variability, but this variation appears to be important in the virus life cycle.

We inferred phylogenetic trees for each of the eight individual gene segments, using our 2196 genes and 2143 full or nearly full-length genes retrieved from GenBank (fig. S1). We observed eight novel clades; two in PB1, one in PB2, two in PA, and three in NP genes (blue brackets). These constitute completely new North American clades that are distinct from

**Table 1.** Project summary. All 11 transcripts from the eight gene segments of AIV are listed along with their lengths in nucleotides (nt), the number of complete gene sequences produced in this study, and the total number of finished nucleotides by segment. The nonsynonymous/synonymous substitution rate ratio (dN/dS), transition/transversion rate ratio (ts/tv), variability of each segment at the amino acid level expressed as a percentage of total ("variability"), and the number of amino acid variations per position in avian versus non-avian hosts are also shown. The asterisks in the length column indicate that this is an average value, because the genes encoding HA, NA, and NS1 have variable lengths.

| Segment | Length (nt) | Segments completed | Finished sequence (nt) | dN/dS | ts/tv | Variability (%) | Variants/position Avian | Variants/position Non-avian |
|---------|-------------|--------------------|-----------------------|-------|-------|-----------------|-------|-----------|
| PB1 | 2341 | 284 | 664,844 | 0.03 | 8.0 | 3.5 | 1.52 | 1.52 |
| PB1-F2 | | | | 9.36 | 33.3 | 10.2 | 3.54 | 2.45 |
| PB2 | 2341 | 272 | 636,752 | 0.03 | 7.9 | 3.7 | 1.67 | 1.66 |
| PA | 2233 | 272 | 607,376 | 0.04 | 7.3 | 3.7 | 1.74 | 1.70 |
| HA | 1770* | 219 | 387,630 | 0.14 | 2.7 | 28.0 | 5.52 | 4.42 |
| NP | 1565 | 292 | 456,980 | 0.03 | 6.9 | 3.7 | 1.81 | 1.85 |
| NA | 1450* | 259 | 375,550 | 0.14 | 3.5 | 28.0 | 5.09 | 4.38 |
| M1 | 1027 | 298 | 306,046 | 0.03 | 8.1 | 3.4 | 1.84 | 1.77 |
| M2 | | | | 0.39 | 12.5 | 4.2 | 3.50 | 3.42 |
| NS1 | 890* | 300 | 267,000 | 0.16 | 5.5 | 6.5 | 3.40 | 2.89 |
| NS2 | | | | 0.12 | 5.9 | 4.7 | 2.81 | 2.68 |
| Total | 13617 | 2196 | 3,702,178 | | | | | |

Asian isolates in the trees. The new surface glycoprotein HA and NA sequences merge with existing clades as defined by serotype; no new clades/serotypes were observed. Likewise, no novel clades were observed for M. The NS tree clearly retains two primary clades (A and B) as previously published (10); however, we observed that the viruses in an emerging branch of the A clade are the same viruses present in branches of clades in other gene segments (boxed in red). This is a family of gull and shorebird viruses (Ciconiiformes) that share genes for NS, M2, NP, PA, and PB2 in our trees. We anticipate that this family, and others found to share genes, will be valuable in correlating genotype with phenotype. To identify other viruses that share genes but do not present obvious branches on a DNA tree, we developed a simple method to visualize unique amino acid signatures.

**Persistence and compensatory mutations revealed by proteotyping.** Traditionally one would assign numbers to all clades in the phylogenetic analysis of individual gene segments and use those numbers to represent and compare genotypes across multiple viruses. In our experience, this approach does not provide the granularity one needs to distinguish subtle, though probably important, differences among these viral gene products. We introduce the concept of proteotyping, in which we identify and number unique amino acid signatures (proteotypes) for sequences that may or may not be distinguished by branches on a phylogenetic tree. We align all sequences of a given segment, manually curate the alignments, generate a maximum likelihood (ML) tree, and then re-sort the alignment to match the order in the tree. We produce an image of the clade-guided sequence alignment by assigning a unique color to each amino acid. Finally, we calculate a consensus sequence for the alignment, hide all the consensus amino acids (colored white to match the background), and remove completely invariant

residue positions. By our method, a residue does not have to occur in the majority of sequences to be the consensus; it only has to occur more than any other residue (6). This leaves only those amino acids that uniquely define a proteotype. Figure 1 illustrates this process for a small portion of the NS tree. From left to right, we illustrate part of the ML tree's clade A with a region of the aligned NS1 sequences and their assigned proteotype numbers. One can clearly identify amino acid signatures that are distinct from consensus despite being grouped in the same tree clade (p1.1 to p1.4). Potential proteotypes were excluded if the sequences were identical or the isolates were consecutive samples from the same location and year. The gull family of viruses, found by chance to share genes in our phylogenetic analysis, is proteotype NSp1.1 and clearly has a distinct amino acid signature. We assigned clade and proteotype numbers to all visually distinct amino acid signatures for all eight gene segments (figs. S2 to S9). Even the highly variable genes encoding HA and NA can be classified in this manner; for example, H6 hemagglutinins clearly have six distinct proteotypes within the H6 clade (fig. S6). Figure S10 presents the compiled proteotype data for all AIV complete genomes, where each column represents a specific gene product in the order PB1, PB2, PA, HA, NP, NA, M, and NS. Each proteotype is assigned a unique color to facilitate the identification of patterns. Using this approach has allowed us to observe for the first time viruses sharing not only specific genes but genes coding for specific proteotypes.

If one sorts the proteotype table by NS type (Fig. 2A), one can easily identify the gull family of viruses we identified first in our phylogenetic analysis (NSp1.1), in addition to two families not previously observed (NSp1.4 and NSp1.5). The NS, M, NP, and PB2 proteotypes in the NSp1.1 family are rare and always occur together, suggesting functionally important

cosegregation. The NSp1.4 family shares HA, NA, M, and NS proteotypes, whereas the NSp1.5 family shares PB1, PA, M, and NS proteotypes. Each family includes viruses isolated from multiple years and each includes proteins with novel signatures pairing with each other. It appears that specific combinations of proteotypes persist over time and may be needed to assemble functional complexes or to maintain specific functional interactions. Sorting this data by each of the remaining seven genes reveals many other instances of virus families sharing specific proteotypes. We anticipate that this method will be useful in identifying AIV protein interactions occurring during infection.

The exchange of gene segments among influenza viruses (reassortment) is one of the hallmarks of influenza virus variation, and these events are rampant in this population. One can see evidence of this in family NSp1.5 in Fig. 2A. Over

**Table 2.** Distribution of PL motifs in 1196 influenza NS1 proteins. PDZ-domain ligand sequences found in this population are listed in the PL column; the top blank entry refers to sequences in which the C terminus is truncated to varying extents. The distribution of each PL sequence in avian, human, swine, and equine AIV isolates is shown. The residue at the −3 position from COOH clearly distinguishes these populations, avian and equine always having E/G at −3 and human almost never having E/G at −3. In contrast, human signatures have R/K at −3 92% of the time; the remaining 8% have avian signatures and are known to be of avian origin. Swine isolates can accommodate both PL motifs. The single isolate with the PL motif KSEV is from the 1918 NS1 sequence. Accession numbers for the public sequences used in this analysis are listed in table S4. Numbers in bold indicate the most frequent motif in avian, human, and swine viruses.

| PL | Avian | Human | Swine | Equine | −3 |
|---|---|---|---|---|---|
|  | 5 | 2 | 4 |  |  |
| EPEV | 23 | 21 | 1 | 14 |  |
| EPKV |  |  |  | 1 |  |
| ESEI | 48 |  | 4 |  |  |
| ESEV | **483** | 12 | 7 | 10 |  |
| ESKV | 55 |  |  |  | E/G |
| GPEV | 2 | 1 | **19** | 2 |  |
| GPKV |  | 6 |  |  |  |
| GSEA | 1 |  |  |  |  |
| GSEI |  |  | 1 |  |  |
| GSEV | 4 |  | 4 |  |  |
| GSKV | 1 | 2 |  |  |  |
| KSEV |  | 1 |  |  |  |
| RPKV |  | 1 |  |  |  |
| RSEA |  |  | 1 |  | R/K |
| RSEV |  | 22 | 4 |  |  |
| RSKI |  | 9 |  |  |  |
| RSKV |  | **403** | **21** |  |  |
| TSEV |  | 1 |  |  |  |
| Totals | 622 | 474 | 73 | 27 |  |
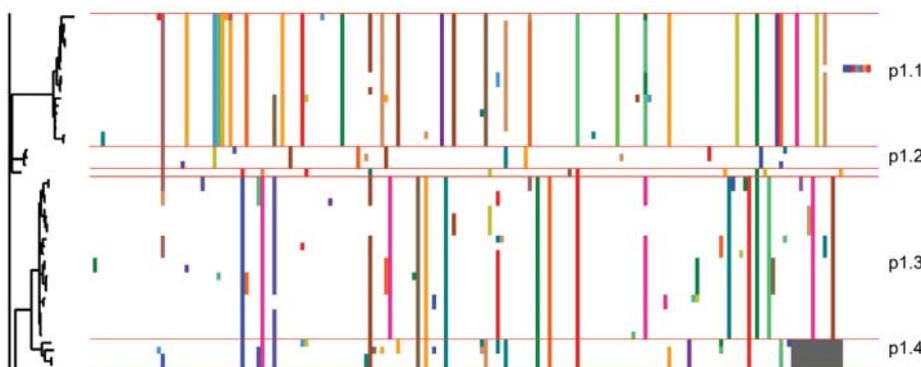


**Fig. 1.** Proteotype assignment for the NS gene/protein. Part of the maximum-likelihood tree within NS's clade A is shown on the left. The amino acid signatures are in the center, and proteotype assignments are shown on the right. White regions in each sequence match the consensus. Proteotypes appear as aligned vertical colored bars, are numbered in order from top to bottom, and are separated by horizontal red lines. The proteotype assignments for the complete NS gene are shown in fig. S2. Clade and proteotype assignments for all eight gene segments are shown in figs. S2 to S9.

the period 1986–1989 the core genes in this family (all those except the surface genes HA and NA) have remained fairly constant, with only a single change in PB2 and NP proteotypes. In contrast, these viruses have acquired four different HAs and six different NAs during this same period.

The data in fig. S10 were re-sorted by HA type, and we illustrate in Fig. 2B the genotype (left) and proteotype (right) for the H6 serotype viruses. The genotyping data show only that these isolates share the H6 serotype/clade HA and M clade 2.0. In contrast, the proteotyping data shows that these viruses can be subclassed into four HA types and two M types and that Mp2.2 is found to pair only with HAp6.2. We propose that the specific pairing we observe in this example results from a change in HA or M requiring the selection of genes encoding proteins with compensating mutations. A num-ber of genetic studies have indicated functional cooperativity between the three polymerase proteins (PB1, PB2, and PA), between NP and M, and between HA and NA (11). Such rela-tionships are clearly observed in our data using proteotypes. The method of assigning proteo-types introduced here has advantages over using traditional clade assignment and phylogenetic analysis alone. Proteotypes provide a higher-resolution view of protein variability, potential protein interactions, and the identity of residues that are likely to be functionally important.
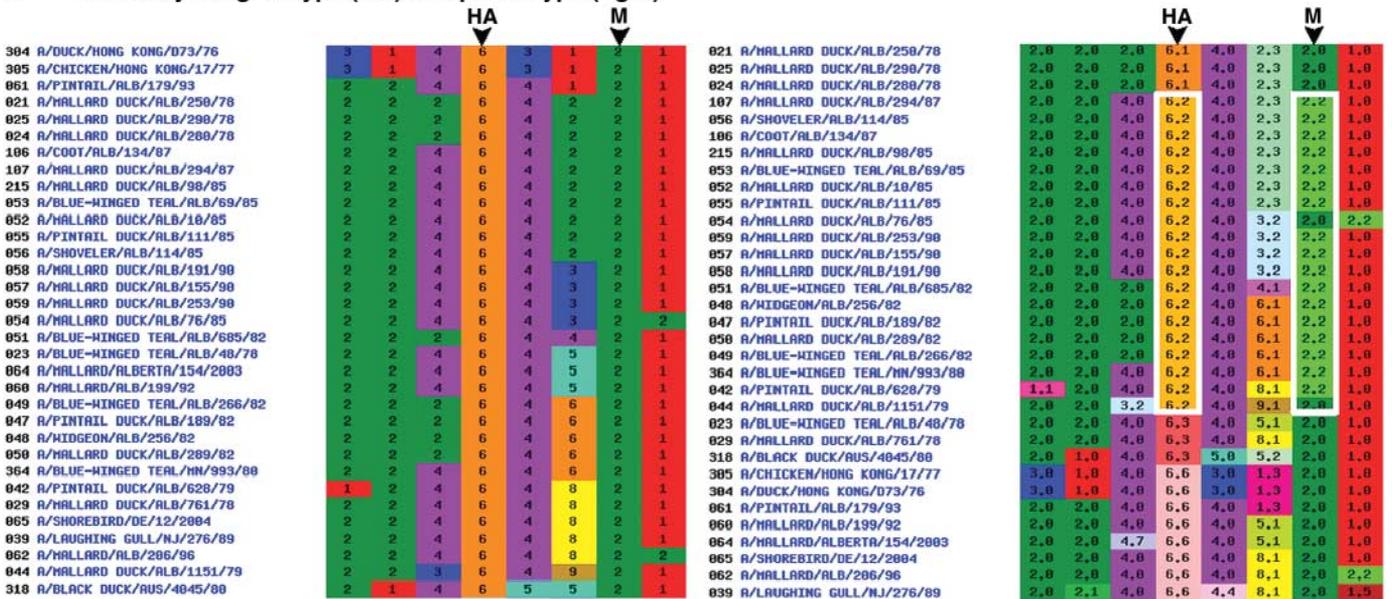
Overall, we are struck by the high frequency of reassortment among the surface glycoproteins and, at the same time, the specific pairing and the conservation seen among the core proteins of AIV. 167 of our 169 complete genomes share five or more genes with at least two additional viruses in our sampling. This is not a sampling artifact, because only four of these isolates have identical genomes and presumably represent multiple sampling of the same virus; the rest clearly share genes with other isolates. The virus families we observed tend to have multiple core genes that specifically pair at the proteotype level, indicating that mutations in these highly conserved genes may be complemented by corresponding mutations in one or more binding or interaction partners in order to result in viable virus. The surface glycoproteins appear to be more freely exchanged than core proteins, most likely because of immune pressure. In addition, any instances of homologous recombination should be easily seen in our proteotypes as contiguous residues that are different from the consensus, but at least among the more conserved internal genes, we have not observed such instances in this preliminary analysis. Our



**Fig. 2.** Proteotyping complete genomes. (**A**) Magnified view of a portion of fig. S10 illustrating three NS families, each sharing different combinations of proteotypes. (**B**) Proteotype data in fig. S10 sorted by HA genotype (left) and proteotype (right). HA and M genotypes and proteotypes are marked to illustrate the detail provided by proteotyping and the recognition that Mp2.2 pairs only with HAp6.2 (white boxes), a feature not seen at the nucleotide level.

findings suggest that homologous recombination is a rare occurrence in AIV evolution.

**PDZ ligand motif in NS1 as a potential virulence determinant.** The contribution of NS to the genetic variability of AIV and the discovery of NS families with unique prototype combinations prompted us to examine this gene in detail. The NS gene encodes the proteins NS1 and NS2. NS1 is found only in infected cells and regulates many cell functions during infection (12). Substantial evidence suggests that NS1 plays a role in virulence by abrogating the expression of antiviral genes in host cells, including interferon (IFN), nuclear factor kappa B (NF-κB), and RNA-activated protein kinase (PKR) pathways (13–17). We examined our collection of new AIV NS data and identified a previously unrecognized canonical PDZ domain ligand (PL) (18) at the C terminus of NS1 (ESEV-COOH). PDZ domains are modular protein interaction domains that bind in sequence-specific fashion to short C-terminal peptides. They are known to function as scaffold proteins that coordinate the assembly of supra-molecular complexes that perform localized signaling at particular subcellular locations (19). Proteins that contain PDZ domains play important roles in many key signaling pathways, including regulating the activity and trafficking of membrane proteins, maintaining cell polarity and morphology, and organizing the postsynaptic density in neuronal cells.

We examined the distribution of the four C-terminal amino acids of 1196 NS1 sequences from avian, human, swine, and equine isolates (Table 2). The PL motif contains E/G (20) at the –3 position in 100% of all 622 available avian and 27 equine NS sequences. In contrast, 436 of 474 human NS1 sequences (92%) have R/K at the –3 position of this motif, and this motif is not seen in avian or equine isolates. Isolates from swine, which can host both avian and human influenza viruses, show instances of both motifs. The sources of the 38 (8%) human isolates with "avian" PL motifs are predominantly H5N1 Asian isolates known to be of avian origin. The human viruses with the EPEV motif (n = 21) are highly virulent isolates from the 1997–1999 outbreaks in Hong Kong. The human isolates with the PL motif ESEV (n = 12) are from the 2003–2004 outbreak in Hong Kong, Vietnam, and Thailand. The 1918 virus NS1 protein has a unique PL motif, KSEV. Thus, the recent high-mortality H5N1 outbreaks are all characterized by NS1 proteins with "avian" PL motifs, whereas previous low-mortality human pandemics (in 1957 and 1968) are characterized by NS1 proteins with "human" PL motifs.

To determine whether these NS1 PL motifs actually bind to PDZ domains and could potentially disrupt key cell pathways, we analyzed the interactions between three known PDZ domains and four synthetic peptides: the two characteristically avian PL motifs found in highly pathogenic human infections in 1997 (EPEV) and 2003 (ESEV); the unique 1918 pandemic virus motif (KSEV); and the predominant motif in low-pathogenic human infections (RSKV). Peptides were produced to contain each motif at the C terminus but were extended by two amino acids (G and S) at the N terminus to increase their length without interfering with binding specificity. Interactions were measured by chemical-shift perturbation nuclear magnetic resonance (NMR) spectroscopy as described by Wuthrich (21) and Wong (22). The results are summarized in Table 3 (NMR data are shown in figs. S11 to S14). The highly pathogenic virus sequences from 2003, 1997, and 1918 bind to the PDZ domain from the protein Disheveled (Dsh), whereas the low-pathogenic human virus sequence (GSRSKV) does not. Similarly, the motif representing low-pathogenic strains shows little or no binding to the first PDZ domain in postsynaptic density protein 95 (PSD-95), but the 2003 and 1918 motifs show strong binding affinities. None of our peptides were found to bind to the seventh PDZ domain in glutamate receptor–interacting protein 1 (GRIP1). Our results suggest that the C terminus of NS1 in low-pathogenic human influenza viruses has no interaction with PDZ domains, but the highly pathogenic viruses from 2003, 1997, and 1918 are all able to interact with some PDZ proteins.

To confirm the peptide results, we expressed histidine-tagged full-length versions of both avian and human NS1 proteins in *Escherichia coli* (Fig. 3A) and assessed their ability to bind to four protein arrays containing 123 human PDZ domains (from TranSignal, Panomics, Fremont, CA). The NS1 protein from A/Blue-winged teal/MN/993/80 (H6N6) contained the predominant avian motif ESEV, and the NS1 protein from A/Memphis/14/98 (H3N2) contained the predominant human motif RSKV. The results demonstrate that the full-length avian NS1 protein
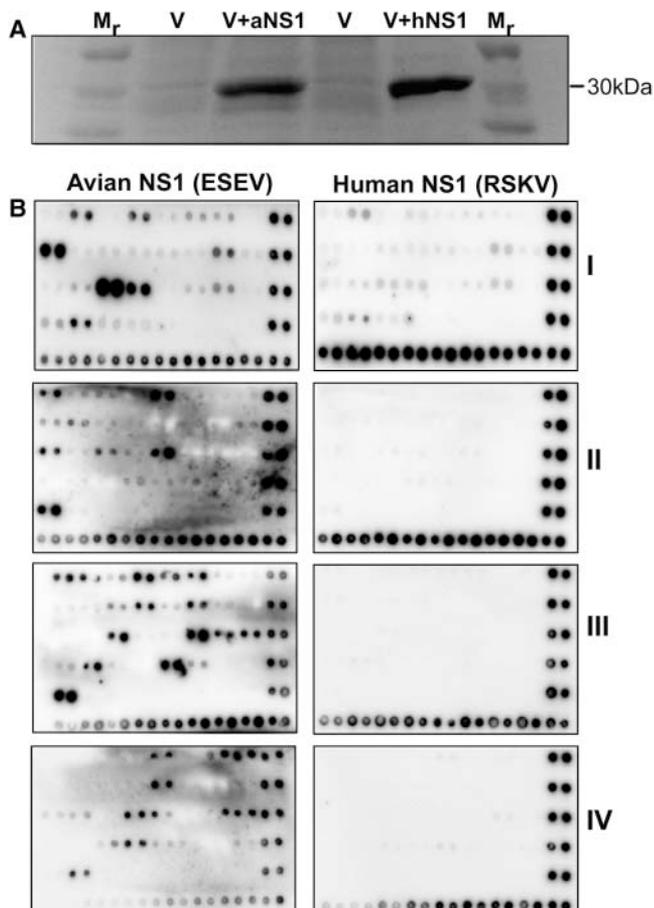


**Fig. 3.** PDZ domain array analysis. (**A**) Avian (V+aNS1) and human (V+hNS1) NS1 clones were expressed as His-tagged proteins in *E. coli* (6). The control lanes are a molecular weight marker (M_r) and vector without insert (V). (**B**) The expressed avian and human NS1 proteins were assayed for their ability to bind 123 human PDZ domains on four membrane filters (I to IV; from Panomics).

**Table 3.** PDZ binding studies. Synthetic peptides corresponding to the two avian PL signatures seen in highly pathogenic human infections in 1997 (GSEPEV) and 2003 (GSESEV), the 1918 pandemic signature (GSKSEV), and the most common low-pathogenic human signature (GSRKSV) were tested for their ability to bind three known human PDZ domains by chemical-shift perturbation NMR. The domains represented are the PDZ domain in Dsh (residues 251 to 345), the first PDZ in PSD-95 (residues 61 to 151), and the seventh PDZ in GRIP1 (residues 980 to 1070). The results are summarized here in terms of relative binding strength. NMR data are shown in figs. S10 to S13.

| | GSESEV (2003) | GSEPEV (1997) | GSKSEV (1918) | GSRSKV (low-pathogenic) |
|---|---|---|---|---|
| Dsh | ++ | ++ | ++ | - |
| PSD-95 | +++ | ± | ++ | ± |
| GRIP1 | - | - | - | - |

binds to 30 different human PDZ domains, whereas the expressed human NS1 protein binds at a very low level or not at all (Fig. 3B). The identities of these 30 PDZ domain–containing proteins are shown in table S3 and include members of all classes of PDZ domain proteins with roles in cell polarity, T cell proliferation, and mitochondrial localization, among others.

Thus, although the molecular consequences of these interactions are as yet unknown, it appears that avian NS1 proteins, when introduced into human cells, have the opportunity to bind to and presumably disrupt many PDZ domain protein–mediated pathways that the human NS1 protein cannot. The 1957 H2N2 and the 1968 H3N2 influenza pandemics were caused by viruses in which only the surface glycoproteins HA and NA and the polymerase protein PB1 of the prevalent human strains were replaced by avian-like molecules, while the remaining core genes remained of human virus origin. In contrast, the recent H5, H7, and H9 outbreaks in Asia were caused by viruses in which the entire complement of influenza genes, including those encoding NS, were derived from an avian source. We propose that the introduction of avian NS1 into human cells can potentially disrupt many cell pathways via binding to PDZ domain–containing proteins, whereas the human NS1 does not. Disruption of these pathways at the cellular level may well contribute to the higher mortality rates reported in the recent outbreaks as compared to those seen in previous pandemics, though it is clear that multiple genes and gene products are involved. This finding reveals an entirely new means by which AIV may interact with host cell proteins, and these proteins may prove valuable as targets for antiviral therapy.

The wealth of AIV genome data provided by this sequencing project has revealed virus families showing conserved combinations of core proteins; frequent reassortment among the surface proteins; newly observed clades of the PB1, PB2, PA, and NP genes; and a possible virulence marker in NS1. We expect that further analysis of this data by the research community will be valuable in understanding AIVs and how they contribute to human disease.

### References and Notes
1. T. Horimoto, Y. Kawaoka, *Nat. Rev. Microbiol.* **3**, 591 (2005).
2. The World Health Organization maintains an updated Web site of the human H5N1 cases and deaths at www.who.int/csr/disease/avian_influenza/country. The figures given here are from the 5 January 2006 update.
3. J. Liu *et al.*, *Science* **309**, 1206 (2005).
4. H. Chen *et al.*, *Nature* **436**, 191 (2005).
5. M. Enserink, *Science* **310**, 209 (2005).
6. Information on materials and methods is available as supporting material on *Science* Online.
7. P. Richterich, *Genome Res.* **8**, 251 (1998).
8. Z. Yang, *Comput. Appl. Biosci.* **13**, 555 (1997).
9. W. Chen *et al.*, *Nat. Med.* **7**, 1306 (2001).
10. Y. Kawaoka *et al.*, *Virus Res.* **55**, 143 (1998).
11. B. W. J. Mahy, in *Genetics of Influenza Viruses*, D. W. Kingsbury, P. Palese, Eds. (Springer-Verlag, Berlin, 1983), pp. 192–254.
12. R. M. Krug *et al.*, *Virology* **309**, 181 (2003).
13. S. Schultz-Cherry *et al.*, *J. Virol.* **75**, 7875 (2001).
14. A. Garcia-Sastre *et al.*, *Virology* **252**, 324 (1998).
15. S. H. Seo, R. G. Webster, *Virus Res.* **103**, 107 (2004).
16. G. K. Geiss *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10736 (2002).
17. A. S. Lipatov *et al.*, *J. Gen. Virol.* **86**, 1121 (2005).
18. Z. Songyang *et al.*, *Science* **275**, 73 (1997).
19. M. Sheng, C. Sala, *Annu. Rev. Neurosci.* **24**, 1 (2001).
20. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
21. K. Wuthrich, *Nat. Struct. Biol.* **7**, 188 (2000).
22. H.-C. Wong *et al.*, *Mol. Cell* **12**, 1251 (2003).
23. We thank the following Hartwell Center staff for superb support: J. Armstrong, S. Tate, and C. Aldridge (HT Sequencing); M. Sanyang (Software Development); S. Olsen, P. Rodrigues, and B. Cassell (Macromolecular Synthesis); and S. Malone and B. Pappas (Operations). Sequences from this study have been deposited in GenBank under accession numbers CY003847 to CY006042. This work was supported by the American Lebanese Syrian Associated Charities, a Cancer Center Support Grant (CA 21765), the U.S. Public Health Service (grant AI95357), and the Hartwell Foundation.

# REPORTS

# Fuel-Powered Artificial Muscles

Von Howard Ebron,[1] Zhiwei Yang,[1] Daniel J. Seyer,[1] Mikhail E. Kozlov,[1] Jiyoung Oh,[1,2] Hui Xie,[1] Joselito Razal,[1] Lee J. Hall,[1] John P. Ferraris,[1] Alan G. MacDiarmid,[1] Ray H. Baughman[1]*

Artificial muscles and electric motors found in autonomous robots and prosthetic limbs are typically battery-powered, which severely restricts the duration of their performance and can necessitate long inactivity during battery recharge. To help solve these problems, we demonstrated two types of artificial muscles that convert the chemical energy of high–energy-density fuels to mechanical energy. The first type stores electrical charge and uses changes in stored charge for mechanical actuation. In contrast with electrically powered electrochemical muscles, only half of the actuator cycle is electrochemical. The second type of fuel-powered muscle provides a demonstrated actuator stroke and power density comparable to those of natural skeletal muscle and generated stresses that are over a hundred times higher.

Although nature's choice is to chemically power the diverse muscles of her design with a high–energy-density fuel, humankind has largely taken another route. In those systems, electrical energy is typically converted to mechanical energy by means of motors, hydraulic systems, or piezoelectric, electrostrictive, or electrochemical actuators (1–9). Because of high electrical power needs, some of the most athletically capable robots cannot freely prance around because they are wired to a stationary power source.

There are exceptions to this use of electrically powered actuators: Chemically powered artificial muscles based on polymer gels were demonstrated over 50 years ago and remain of practical interest for both chemically and electrically powered actuators (10–12). Although actuator strains can be very large, their application has been limited by low response rates, low stress generation, and the low energy densities of the chemicals used for driving actuation. The combustion of fuels in a preburner has been used to indirectly power actuation of shape-memory alloys (13), and muscles that act as fuel cells have been proposed (14, 15) but not experimentally demonstrated. Also, nanoscale and larger actuators that are powered by oxygen gas released by the catalytic decomposition of hydrogen peroxide have been described (16–20).

We experimentally demonstrated two types of artificial muscles that are powered by high–energy-density fuels (hydrogen, methanol, or formic acid). The first type uses a catalyst-containing carbon nanotube electrode that simultaneously functions as a muscle, a fuel-cell electrode, and a supercapacitor electrode. The result is a muscle that converts chemical energy in a fuel to electrical energy and can use this electrical energy for actuation, store it, or potentially use it for other energy needs. The

[1]Department of Chemistry and NanoTech Institute, University of Texas at Dallas, Richardson, TX 75083–0688, USA. [2]Research Center of Dielectric and Advanced Matter Physics and Department of Physics, Pusan National University, Busan 609–735, Korea.

*To whom correspondence should be addressed. E-mail: ray.baughman@utdallas.edu