Schein et al. BMC Bioinformatics 2012, 13(Suppl 13):S9
http://www.biomedcentral.com/1471-2105/13/S13/S9

BMC Bioinformatics

RESEARCH Open Access

# Physicochemical property consensus sequences for functional analysis, design of multivalent antigens and targeted antivirals

Catherine H Schein[1,2,3*], David M Bowen[1,2], Jessica A Lewis[4], Kyung Choi[1,2], Aniko Paul[5], Gerbrand J van der Heden van Noort[6], Wenzhe Lu[2], Dmitri V Filippov[6]

*From* The 8th Annual Biotechnology and Bioinformatics Symposium (BIOT-2011)
Houston, TX, USA. 20-21 October 2010

## Abstract

**Background:** Analysis of large sets of biological sequence data from related strains or organisms is complicated by superficial redundancy in the set, which may contain many members that are identical except at one or two positions. Thus a new method, based on deriving physicochemical property (PCP)-consensus sequences, was tested for its ability to generate reference sequences and distinguish functionally significant changes from background variability.

**Methods:** The PCP consensus program was used to automatically derive consensus sequences starting from sequence alignments of proteins from Flaviviruses (from the Flavitrack database) and human enteroviruses, using a five dimensional set of Eigenvectors that summarize over 200 different scalar values for the PCPs of the amino acids. A PCP-consensus protein of a Dengue virus envelope protein was produced recombinantly and tested for its ability to bind antibodies to strains using ELISA.

**Results:** PCP-consensus sequences of the flavivirus family could be used to classify them into five discrete groups and distinguish areas of the envelope proteins that correlate with host specificity and disease type. A multivalent Dengue virus antigen was designed and shown to bind antibodies against all four DENV types. A consensus enteroviral VPg protein had the same distinctive high pKa as wild type proteins and was recognized by two different polymerases.

**Conclusions:** The process for deriving PCP-consensus sequences for any group of aligned similar sequences, has been validated for sequences with up to 50% diversity. Ongoing projects have shown that the method identifies residues that significantly alter PCPs at a given position, and might thus cause changes in function or immunogenicity. Other potential applications include deriving target proteins for drug design and diagnostic kits.

## Background

The most useful information one can glean from aligned sequences of proteins is first, the absolutely conserved residues, which are usually those that maintain the structure of the protein or are vital for function. The pattern, or profile of conserved residues in an alignment of a protein type can be used to identify proteins in the same group, that may have a similar structure [1-4]. In addition, the variance within the sequences, which may occur at specific positions due to random variation (i.e., in RNA viruses, an error prone polymerase), can also indicate a functional change. It is thus important to be able to separate background variation, which, in our approach, is assumed to cause little change in the physicochemical properties (PCPs) at a position, from those that alter these properties sufficiently to lead to a variance in protein function or immunogenicity [5-8].

* Correspondence: chschein@utmb.edu
[1]Institute for Translational Sciences, Computational Biology, Sealy Center for Structural Biology and Molecular Biophysics, University of Texas Medical Branch, Texas 77555–0857, USA
Full list of author information is available at the end of the article

BioMed Central
© 2012 Schein et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Very large alignments present intrinsic problems in discriminating residue conservation or patterns of variance, and require special software even to view them. Another problem in dealing with biological datasets, such as the many Flavivirus sequences we have collected within the Flavitrack database [9,10], is that they often have a pronounced bias due to unequal distribution, which can arise from non-uniform sampling [11]. For example, one may have many closely related sequences from one epidemic, where serious infections occurred, but few from the intervening years, when most infections had a less lethal phenotype. Conventional methods for calculating consensus sequences assume an unbiased data set, and typically calculate only the most common amino acid in a column [12]. An example of such a consensus (Figure 1A) shows that while it provides useful information on the degree of conservation of the amino acids in aligned sequences, it cannot suggest a rational choice of amino acid at highly variant positions. Profiling methods [13-15] based on amino acid scoring matrices can also be used to obtain a consensus sequence, but these are primarily designed to detect distantly related members of a set of proteins.

Here we show applications of a general method to calculate physicochemical property (PCP)-consensus sequences. The method is designed to filter noise due to random amino-acid variations within strains or subtypes from more significant variation. We first modeled PCP-consensus sequences for several proteins, and showed that they were stable after minimization with our FANTOM program. We have also produced several PCP-consensus proteins from synthetic gene sequences in *E. coli* and tested their ability to be recognized enzymatically and immunologically. As discussed below, PCP-consensus sequences have many uses, in sequence classification, epitope comparison, in defining multivalent sequences as immunogens for vaccine use, and for defining targets for multivalent drug design.

## Methods

### Deriving PCP consensus sequences

Our method assumes that one has a high quality alignment, of any number $N$ of sequences with a maximum length L. Choosing an appropriate sequence grouping is a chicken/ egg problem, and will be discussed in more detail below. Multiple sequence alignments were

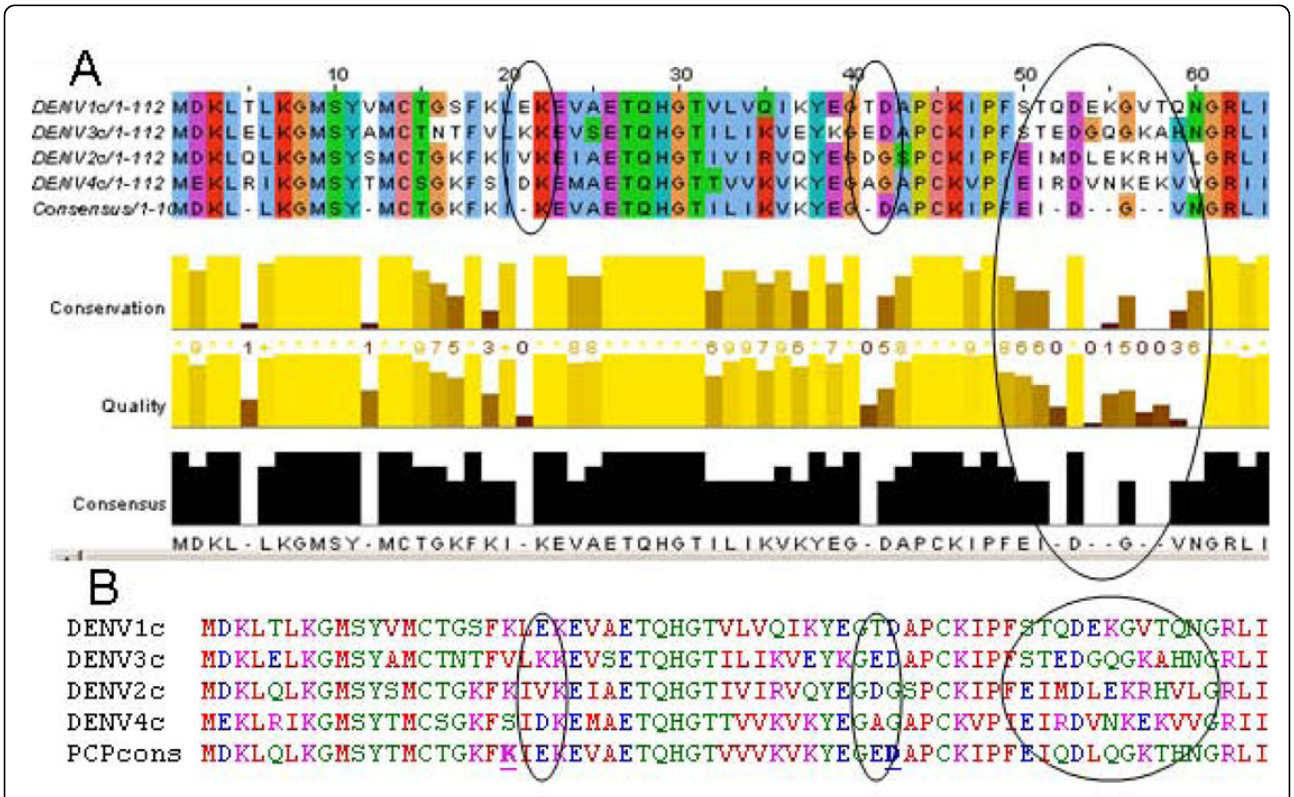


**Figure 1 Dealing with variable positions in aligned sequences.** A consensus that chooses only the most frequently occurring amino acid at a given column of an alignment (e.g., that from the Jalview applet of Clustal W) is best for indicating highly conserved residues, but does not give any consensus value for variable positions (such as those circled). B) A PCP-consensus for the same sequences selects the amino acid that is most similar in properties to all others in the set.

generated with Clustalw 2.0.3[16], or MUSCLE [17,18], for very large alignments, using default parameters. It is best to check such large alignments for inappropriate gapping. Although there are statistical methods to do this [19], for the purposes of this early validation work, we have chosen homologous proteins where a representative protein structure is known, and sequence groupings that have more than 50% identity. This allows us to check that any gapping is consistent with secondary structure elements, and conservation of disulfide bonds and salt bridges.

Figure 2 compares the differences in the purpose and method for obtaining a PCP consensus sequence with those of profiling methods. To obtain PCP-consensus sequences, at each position of the multiple alignment, one amino acid is chosen that best approximates the average value of the PCPs. The PCPs of the 20 amino acids are defined by a set of numerical descriptors, 5 Eigenvectors obtained by multidimensional scaling of over 200 unique property measurements [20]. Each amino acid can be discriminated from all the others as points in a five dimensional space, where the five dimensions, the first 5 eigenvectors, roughly correspond to hydrophobicity/hydrophilicity (E1); size (E2); alpha-helix propensity (E3); the property E4 is related to the partial specific volume, number of codons and relative abundance of the amino acids; and E5 correlates weakly with beta-strand propensity [20]. This 5-dimensional approach to similarity allows one to calculate a true consensus amino acid, i.e., the one closest in its PCPs to all others in the column, even at very variable positions. Given an amino acid alignment, the program selects an amino acid that is closest in the property space to the average values for all the other amino acids at each position. First, the average value of each of the 5 PCP-vectors $p$=E1,..,E5 is determined at each column, effectively turning an alignment of $N$ sequences of maximum length L into a 5xL matrix:

$$\bar{E}^p = \sum_{j=1}^{N} \frac{V_j^p}{N}, \tag{1}$$

where $N$ is the number of amino acids in the given column of the alignment; and $V_j^p$, the value for the relevant p eigenvector for the jth amino acid. Then the consensus amino acid ($A_a$) is chosen from those occurring naturally at that position with the least Euclidean distance from the average:

$$\left| A_a, \bar{E} \right| = \sqrt{\sum_{p=E1}^{E5} b_p \left| V_a^p - \bar{E}^p \right|^2}. \tag{2}$$

The alignment independent scale factors $b_p$ were calculated so that vector values with higher relative entropies at a given column would be more significant, and were calculated as described elsewhere [9].

For very variable positions or highly biased datasets, the amino acids that naturally occur at each position can be used one time, without regard to their rate of

| PCP-Consensus | Profile Methods |
|---|---|
| Finds consensus AA based on 5 PCP-Eigenvectors | Finds consensus AA based on simply abundance, or adjusted according to substitution matrices |
| Identifies structural and functional variability in highly related sequences (e.g. same viral family) | Identifies structural commonalities even among diverse sequences (E > 0.001) |
| Determines a single sequence closest in PCPs to all others in the set | Can summarize the most conserved properties in the form of a numerical matrix |

5

**Figure 2** Comparing the PCP-consensus method and its uses to profiling methods and their applications.

occurrence in the column, to calculate the average values of the 5 property vectors. In that case, equation 2 can still be used, and the chosen "consensus" amino acid is simply that closest in its physical properties to all the naturally occurring amino acids. Other possibilities for dealing with bias in the data set, such as selective sequence weighting can also be used [21] to determine the property averaging method. This is an area for further study, as a completely mathematical solution for all situations is probably not possible.

It should be stressed that biological findings can be incorporated at any point in this process, in distinguishing sequences that have specific properties. Bioinformaticians should be aware that sequences grouped according to a biological assay may or may not correlate with distinctive genotypes. For example, the four types of Dengue viruses (DENV1-4), first characterized by Sabin in the early fifties based on immunological reactivity [22], segregate rather cleanly into four distinct genotypes (see below). However, human enteroviruses (HEV), designated Coxsackie virus A or B based on the type of paralysis they caused in newborn mice, did not separate neatly into two distinct sequence groups [23]. While we have relied on the strain designations in the NCBI for Flaviviruses, other useful data that should be part of the functional annotation (such as lethality) is often not specified by those providing the sequences to NCBI.

**Models of PCP-consensus sequences** were prepared with our MPACK modeling suite [24-27] using the crystal structure of the DENV-2 protein (1OAN.pdb) [28].

## Results and discussion

### PCP-consensus sequences using Flavitrack entries

There are many potential uses for PCP-consensus sequences in virology, for example in classifying strains, identifying functional alterations [29], and in designing novel, multivalent antigens for vaccines and diagnostics. Here we will show some applications based on data stored in our Flavitrack database (http://carnot.utmb.edu/flavitrack), which is a compendium of annotated Flavivirus sequences [9,10]. Flaviviruses (FV), which include yellow fever (YFV), DENV, and West Nile viruses (WNV), are important human and animal pathogens which typically require insect vectors to infect mammalian hosts [30-35]. While mosquito control can be effective, antiviral agents and wide-spectrum vaccines are being sought to protect those in endemic areas [36-43]. To design effective vaccines, the areas of the viral proteins required for virus function or infectivity should be targeted by antibodies. Flaviviruses are variable, with many sequence variants found even in single virus isolates from the same patient, so-called "quasispecies" [44]. However, when catalogued, the strains appear redundant from a mathematical standpoint, with interstrain diversity occurring at fewer than 1% of positions. While much of this variation is neutral for phenotype, even a single point mutation can greatly alter the immunogenicity of the envelope protein or alter virus entry [38,40,45-47]. Recognizing such function-altering amino acid substitutions is important for designing vaccines that will protect against many Flaviviruses simultaneously, and entry inhibitors.

Our first programs for analyzing the sequences in Flavitrack attempted to highlight all variation in the aligned sequences in a fashion suitable for conventional visual scanning of the data. These first attempts illustrated the need for unbiased data reduction: an alignment of 928 sequences (Flavitrack ca. 2009) covered dozens of pages of paper. Even at the smallest possible type (a microtext version of the database sequences provided to us by Reiner Eschbach's group at Xerox), no screen setting was adequate to view more than a small part of the data. Also for the purposes of determining variation, we needed a rational mean sequence to compare sequences. Other intrinsic problems in the data were the non-random sequence distribution, with many more sequences available for certain mosquito-borne viruses (WNV and DENV) than for any of the tick-borne or no-known-vector (NKV) groups.

### PCP-consensus reference strains

The traditional Flavivirus group reference strains are historical isolates with defined immunological properties, often predating the genome era, and not chosen to best represent a consensus genotype. These strains may have been passaged many times in the lab setting, a process that could result in a sequence quite distant from the original or any subsequently isolated wild type strain. Now that direct PCR sequencing from the original isolate is possible, a more relevant methodology would be to use a series of PCP-consensus reference strains, as long as these were shown to correlate with serotype data or other biological assays. We began by creating PCP-consensus sequences for each group of Flaviviruses, where the groups were defined according to the strain designations in the NCBI headers for the annotated sequences in Flavitrack [48]. A series of 37 PCP-consensus strains for the most common groups of Flaviviruses was derived (Figure 3), which allowed comparison of the overall properties of the individual virus types and separated the Flaviviruses, based on two different proteins, into five discrete groupings. We suggest that these can be viewed as reference strains for grouping (or profiling [49]) novel Flavivirus isolates, obtained for example, from sentinel screening of mosquitoes [50].

Proper grouping of virus isolates has meaning beyond mere nomenclature: comparing these consensus

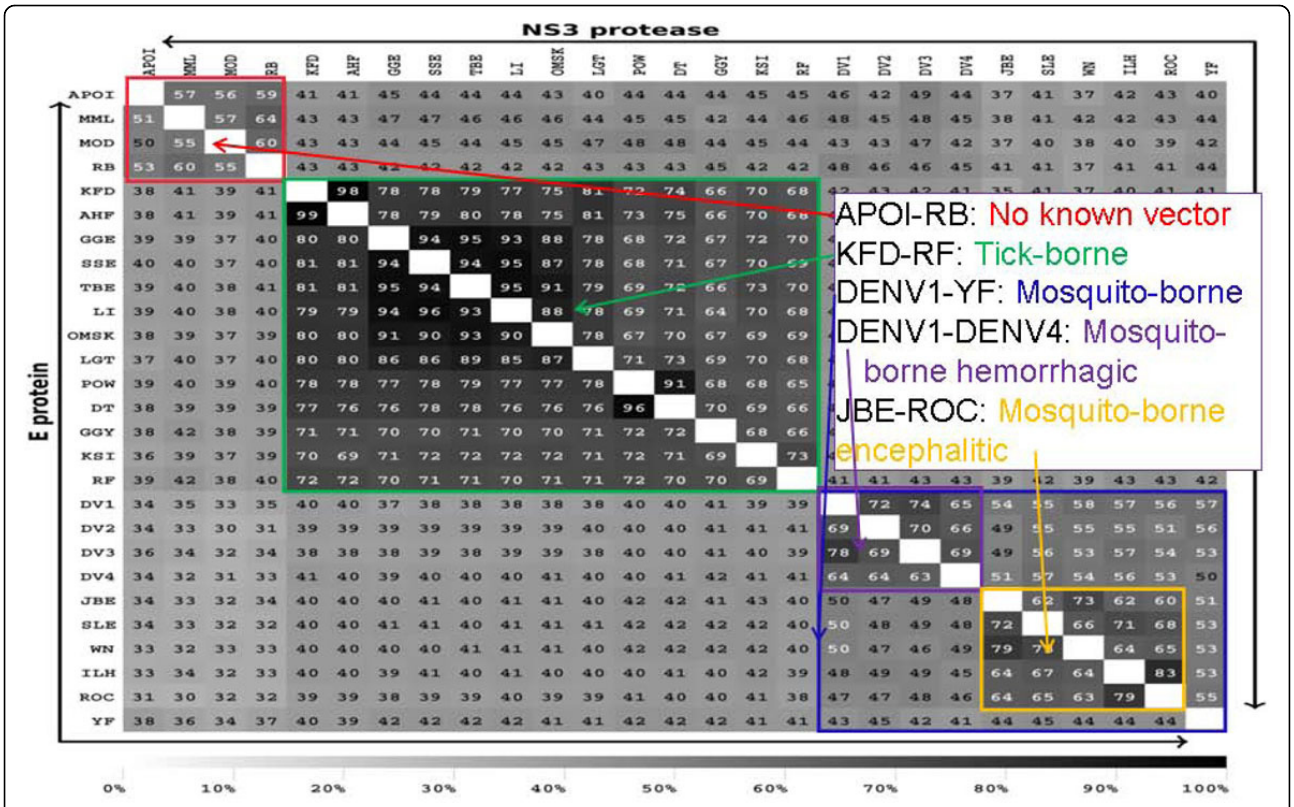


**Figure 3 PCP-consensus sequences distinguish 5 groups of Flaviviruses**. The PCP- consensus sequences of two viral proteins, the envelope (E) and NS3 protease, of 37 different Flavivirus strains derived from 928 annotated entries in Flavitrack [48], separate into five distinct clusters, based on their percent identity to one another.

sequences, we could discriminate residue changes that fell outside the expected group variance. The comparison highlighted insertions and deletions that correlated with whether a species was carried by mosquitoes or ticks, and even with the type of disease (encephalitic vs. haemolytic) resulting from human infection [48]. Additional uses of classification are to detect when a strain that appeared to be adapted to growth only in mosquitoes or bats contained key substitutions that might indicate human cross-over potential.

### A multivalent PCP-consensus DENV antigen

Understanding the variation between viral types is particularly important for DENV, as it has been shown that reinfection of a person carrying antibodies against one DENV immunotype with a different type can result in Dengue hemorrhagic fever (DHF), a severe disease that requires proper medical support. In 2010, an epidemic in Brazil caused over a million documented cases, with about 600 deaths (many of which were young children). Billions of people throughout the world are at risk for DENV [43,51-55]. We approached this problem by generating a PCP-consensus of the representative envelope proteins for the four different types of Dengue virus (DENV1-4). This consensus was close to each type and thus represented a good mean (Figure 4). This was especially clear for the outlier, DENV-4, which was lowest in absolute identity to all the other three DENV types and is the hardest to neutralize with antibodies generated by infection with DENV-1,2 or 3. Figure 5 shows a portion of the consensus envelope protein of Dengue, with residues of maximum variation highlighted. Optimal choice of residues in variable areas of the viral strains can guide the design of multivalent vaccines and inhibitors. Figure 6 shows how this sequence was further developed for testing in animals. The recombinant protein, after optimization to reflect the DENV4 outlier, bound antibodies to all four DENV types [56] and preliminary work has shown that inoculation of this protein generates multivalent antibodies in rabbits and mice that neutralized all four DENV strains (data not shown).

### PCP-consensus VPg for designing multivalent inhibitors of human enteroviruses

To further illustrate the potential uses of the method, we designed and produced a PCP-consensus "viral peptide linked to the genome" (VPg) for the human enteroviruses (HEV), which include polioviruses (PV), Coxsackie

| | DENV1c | DENV3c | DENV2c | DENV4c | PCPcon |
|---|---|---|---|---|---|
| DENV1c | - | 78 | 68 | 63 | 82 |
| DENV3c | 74.1 | - | 68 | 62 | 78 |
| DENV2c | 66.1 | 62.5 | - | 64 | 80 |
| DENV4c | 56.2 | 51.8 | 60.7 | - | 72 |
| PCPcon | 81.2 | 75.9 | 76.8 | 67.9 | - |

**Figure 4 Identity matrix for four individual DENV consensus strains and a multivalent PCP consensus derived from them.** The matrix shows the inter-sequence Clustal W scores (top) or % Identity (bottom) for PCP-consensus sequences for the envelope protein of each of the four DENV types, and a PCP consensus prepared from these four individual consensus sequences. The overall PCP-consensus is about equidistant from the four consensus sequences, using either metric for similarity. Note the four distinct genotypes, that DENV 1 and 3 are closer to each other than to DENV2, and that DENV4's sequence is distant from the other 3 types.

viruses A and B (CVA and CVB), and Echovirus. To initiate RNA synthesis, HEV polymerases (3D-pol) uridylylate a conserved tyrosine residue in the 22 amino acid long VPg to form VPgpU. We have determined the NMR structure of poliovirus type 1 (PV1)-VPg and PV1-VPgpU and shown that uridylylation stabilized the 3D-structure of the peptide, which is probably necessary for VPgpU to serve as a precursor for RNA synthesis [57-59]. As this reaction is not found in normal cells, it is a target for antiviral drug design [57]. To develop a

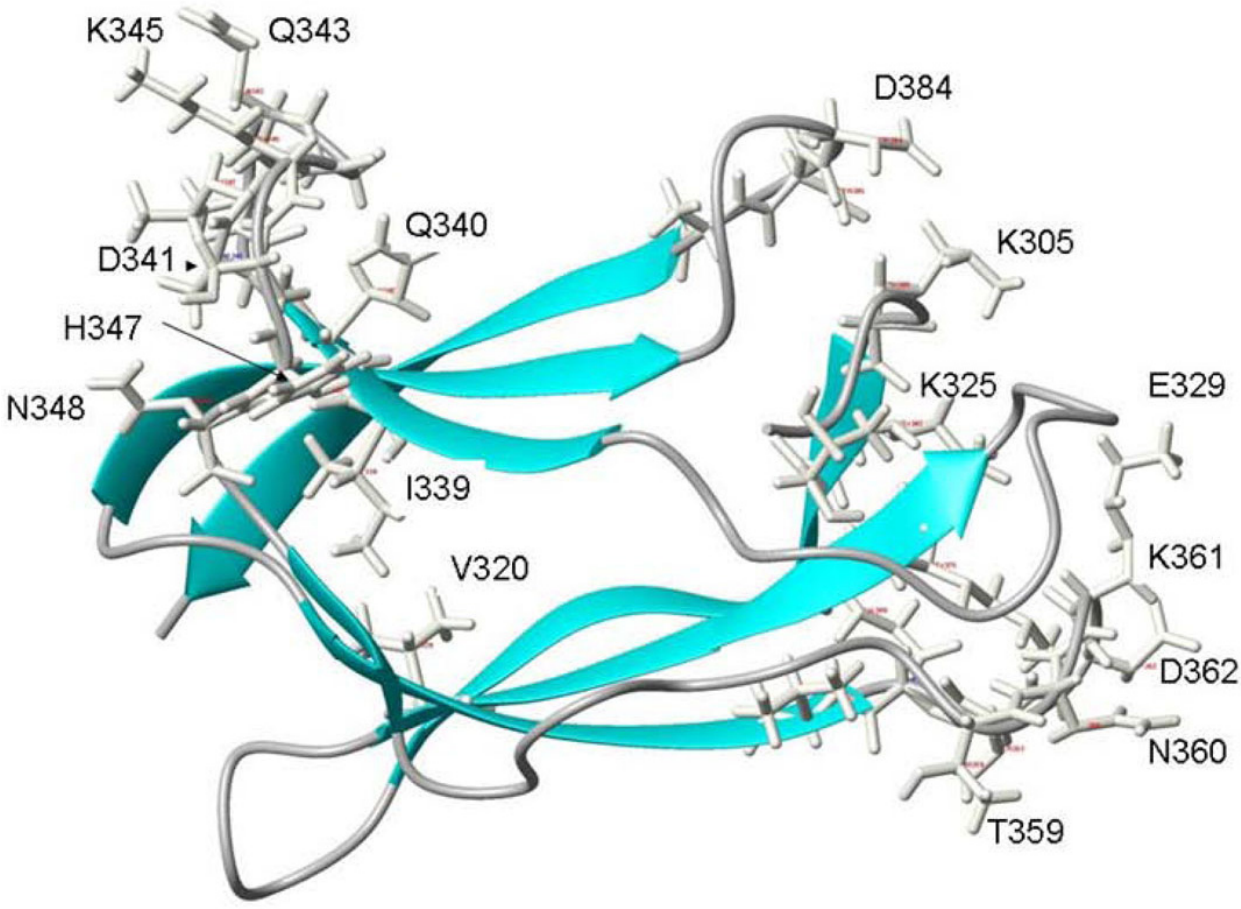


**Figure 5 Variable positions in a PCP-consensus DENV antigen.** A model of a PCP-consensus EdomIII DENV antigen is shown in ribbon format, with the sidechains where maximum variation occurs shown in stick format. Labels on the outermost residues are shown for orientation, and to illustrate that the type specific epitope surfaces are not linearly encoded in the amino acid sequence.

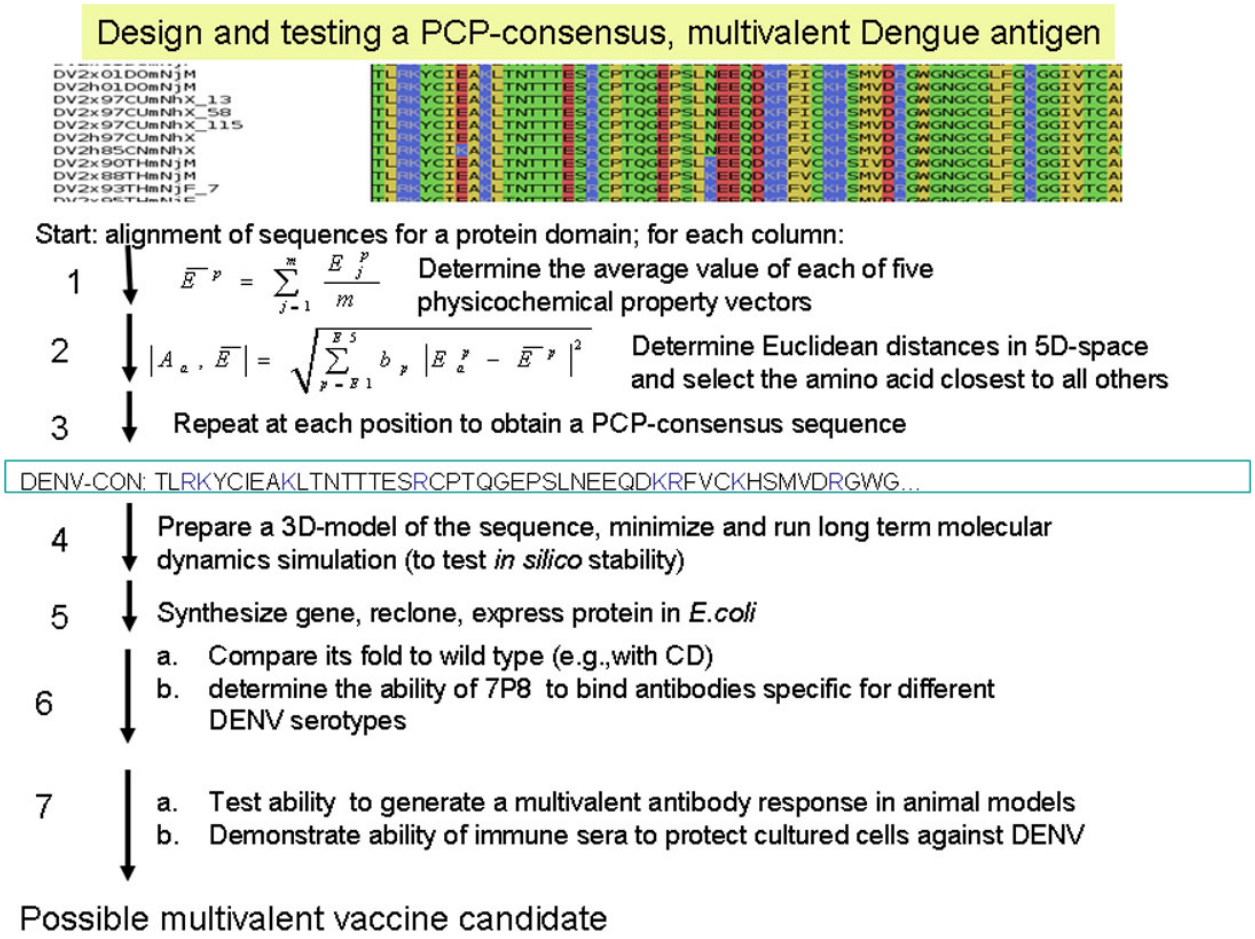


**Figure 6 Deriving a multivalent DENV antigen for testing in animals.** Designing, producing and testing a multivalent DENV antigen for vaccine use. The protein was able to bind antibodies generated against all four types of DENV, and induced antibodies in a rabbit that recognized wild type proteins from all four types of DENV.

multivalent target VPg suitable for designing inhibitors against all HEV, the sequences of 33 unique HEV-VPgs were aligned and a PCP consensus protein, VPg-cons, was prepared with our automatic program. Although only about 50% of the amino acids were conserved completely the selected VPgs, the calculated pKa values of all of them were exactly 10.46, suggesting the peptide must be very basic in order to function (this is consistent with our NMR structures, which shows the basic, essential side chain of Arg17 very close to the phosphates of the coupled UMP). The PCP-consensus VPg, which is not identical to any naturally encoded sequence, had the same calculated pKa of 10.46. This illustrates that the consensus represents conserved physicochemical parameters of a sequence set. Both the PCP-consensus HEV-VPg and HEV-VPgpU were prepared synthetically [60,61]. The HEV-VPg can be uridylylated by both the PV1- and CVA24-RNA-polymerases as well or better than the wild type VPg encoded in their respective genomes. Thus the PCP-consensus VPg represents the conserved properties of HEV wild type VPgs, and functions in a multivalent manner. Further study of the structure of the HEV-VPgpU should aid in deriving a general mechanism for uridylylation. Inhibitors based on the consensus HEV sequence should be multivalent, and prevent replication of all HEVs.

## Conclusions

Defining PCP-consensus sequences can aid in analysis of large sequence datasets. The calculation method, based on a previously validated 5D-vector scale for the physicochemical properties of the amino acids, is straightforward, once a suitable alignment of related sequences is obtained. Having a rational consensus allows one to distinguish residue variations that significantly alter the properties at a given position. The method is thus suitable for application to many types of bioinformatics data.

The usefulness of the methodology in virology was demonstrated in two practical applications. A multivalent, PCP-consensus DENV vaccine candidate was designed, produced, and shown to bind antibodies against all four types of DENV. Also, a consensus HEV-VPg has similar properties, particularly pKa, conserved in wild type VPgs, and was uridylylated by two different HEV polymerases. This validated method should find application in many practical areas of virology and other areas of biology.

**Acknowledgements**
We thank all coworkers from the UTMB, especially David Beasley for his invaluable assistance with all the DENV work, and Werner Braun for his thoughtful input on handling bias and alternate methods of consensus sequence determination; and the Xerox team of Dr. Reiner Eschbach for microtext versions of alignments.
**Funding**: The DENV vaccine project was supported in part by grant 1UL1RR029876-01 from the National Center for Research Resources, NIH to the Institute for Translational Studies of the UTMB, (support was to CHS and David Beasley for pilot protocols #763 and #809). Jessica A. Lewis is supported by a pre-doctoral Fellowship from the Sealy Center for Vaccine Development at the UTMB; Dr. Paul work is supported by NIH AI015122 to E. Wimmer. Kay Choi's work is supported by NIH grant 1R01AI087856. Development of the PCP-consensus method was supported in part by NIH grant AI064913 (to Werner Braun and CHS) and EPA-STAR grant RE-83406601-0 (to CHS). D.V.F. and G.J. v.d. H. v. N. are supported by The Netherlands Organization for Scientific Research (NWO). The computational resources of the Sealy Center for Structural Biology and Molecular Biophysics were also used in this project.
This article has been published as part of *BMC Bioinformatics* Volume 13 Supplement 13, 2012: Selected articles from The 8th Annual Biotechnology and Bioinformatics Symposium (BIOT-2011). The full contents of the supplement are available online at http://www.biomedcentral.com/1471-2105/13/S13/S1

**Author details**
[1]Institute for Translational Sciences, Computational Biology, Sealy Center for Structural Biology and Molecular Biophysics, University of Texas Medical Branch, Texas 77555–0857, USA. [2]Departments of Biochemistry and Molecular Biology, University of Texas Medical Branch, Texas 77555–0857, USA. [3]Microbiology and Immunology, University of Texas Medical Branch, Texas 77555–0857, USA. [4]Department of Pathology, University of Texas Medical Branch, Galveston, Texas 77555–0857, USA. [5]Dept. of Molecular Genetics and Microbiology, Stony Brook University, Stony Brook, NY 11790, USA. [6]Leiden Institute of Chemistry, Leiden University, PO Box 9502, 2300 RA Leiden, The Netherlands.

**Authors' contributions**
CH developed the PCP-consensus method and participated in all the experiments described, as well as writing the paper. DB and JL prepared the PCP-consensus DENV antigens and did ELISA tests for binding. KC prepared polymerases for the VPg assay; AP did assays for uridylylation of the consensus VPg with two polymerases; WL (graduate student) did PCP-consensus calculations; GHN and DF prepared PCP-consensus VPg and VpgpU peptides.

**Competing interests**
The authors declare that they have no competing interests.

Published: 24 August 2012

**References**

1. Przybylski D, Rost B: **Consensus sequences improve PSI-BLAST through mimicking profile-profile alignments.** *Nucleic Acids Res* 2007, **35(7)**:2238-2246.
2. Schein CH, Ozgun N, Izumi T, Braun W: **Total sequence decomposition distinguishes functional modules, "molegos" in apurinic/apyrimidinic endonucleases.** *BMC Bioinformatics* 2002, **3**:37.
3. Schein CH, Zhou B, Braun W: **Stereophysicochemical variability plots highlight conserved antigenic areas in Flaviviruses.** *Virol J* 2005, **2**:40.
4. Schein CH, Zhou B, Oezguen N, Mathura VS, Braun W: **Molego-based definition of the architecture and specificity of metal-binding sites.** *Proteins* 2005, **58(1)**:200-210.
5. Schein CH: **The shape of the messenger: using protein structure information to design novel cytokine-based therapeutics.** *Curr Pharm Des* 2002, **8(24)**:2113-2129.
6. Schein CH: **From interleukin families to glycans: relating cytokine structure to function.** *Curr Pharm Des* 2004, **10(31)**:3853-3855.
7. Schein CH, Haugg M: **Deletions at the C-terminus of interferon gamma reduce RNA binding and activation of double-stranded-RNA cleavage by bovine seminal ribonuclease.** *Biochem J* 1995, **307(Pt 1)**:123-127.
8. Schein CH, Ivanciuc O, Braun W: **Common physical-chemical properties correlate with similar structure of the IgE epitopes of peanut allergens.** *J Agric Food Chem* 2005, **53(22)**:8752-8759.
9. Danecek P, Schein CH: **Flavitrack analysis of the structure and function of West Nile non-structural proteins.** *Int J Bioinform Res Appl* 2010, **6(2)**:134-146.
10. Misra M, Schein CH: **Flavitrack: an annotated database of flavivirus sequences.** *Bioinformatics* 2007, **23(19)**:2645-2647.
11. DeBiasi RL, Tyler KL: **West Nile virus meningoencephalitis.** *NATURE CLINICAL PRACTICE NEUROLOGY* 2006, **2(5)**:264-275.
12. Ramanathan MP, Kutzler MA, Kuo YC, Yan J, Liu H, Shah V, Bawa A, Selling B, Sardesai NY, Kim JJ, *et al*: **Coimmunization with an optimized IL15 plasmid adjuvant enhances humoral immunity via stimulating B cells induced by genetically engineered DNA vaccines expressing consensus JEV and WNV E DIII.** *Vaccine* 2009, **27(32)**:4370-4380.
13. Gribskov M: **Profile analysis.** *Methods Mol Biol* 1994, **25**:247-266.
14. Gribskov M, McLachlan AD, Eisenberg D: **Profile analysis: detection of distantly related proteins.** *Proc Natl Acad Sci U S A* 1987, **84(13)**:4355-4358.
15. Merkeev IV, Mironov AA: **PHOG-BLAST–a new generation tool for fast similarity search of protein families.** *BMC Evol Biol* 2006, **6**:51.
16. Thompson JD, Gibson TJ, Higgins DG: **Multiple sequence alignment using ClustalW and ClustalX.** *Curr Protoc Bioinformatics* 2002, **Chapter 2**:Unit 2-3.
17. Edgar RC: **MUSCLE: a multiple sequence alignment method with reduced time and space complexity.** *Bmc Bioinformatics* 2004, **5**:113.
18. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32(5)**:1792-1797.
19. Bailey TL, Gribskov M: **Estimating and evaluating the statistics of gapped local-alignment scores.** *J Comput Biol* 2002, **9(3)**:575-593.
20. Venkatarajan MS, Braun W: **New quantitative descriptors of amino acids based on multidimensional scaling of a large number of physical-chemical properties.** *Journal of Molecular Modeling* 2001, **7(12)**:445-453.
21. May AC: **Optimal classification of protein sequences and selection of representative sets from multiple alignments: application to homologous families and lessons for structural genomics.** *Protein Eng* 2001, **14(4)**:209-217.
22. Sabin AB: **Research on Dengue during World War II.** *The American Journal of Tropical Medicine and Hygiene* 1952, **1(1)**:30-50.
23. **Molecular Biology of Picronaviruses.** ASM Press;Bert L Semler and Eckard Wimmer 2002:.
24. Ivanciuc O, Oezguen N, Mathura VS, Schein CH, Xu Y, Braun W: **Using property based sequence motifs and 3D modeling to determine structure and functional regions of proteins.** *Curr Med Chem* 2004, **11(5)**:583-593.
25. Oezguen N, Zhou B, Negi SS, Ivanciuc O, Schein CH, Labesse G, Braun W: **Comprehensive 3D-modeling of allergenic proteins and amino acid composition of potential conformational IgE epitopes.** *Mol Immunol* 2008, **45(14)**:3740-3747.
26. Soman KV, Midoro-Horiuti T, Ferreon JC, Goldblum RM, Brooks EG, Kurosky A, Braun W, Schein CH: **Homology modeling and characterization of IgE binding epitopes of mountain cedar allergen Jun a 3.** *Biophys J* 2000, **79(3)**:1601-1609.
27. Soman KV, Schein CH, Zhu H, Braun W: **Homology modeling and simulations of nuclease structures.** *Methods Mol Biol* 2001, **160**:263-286.

28. Modis Y, Ogata S, Clements D, Harrison SC: **Variable surface epitopes in the crystal structure of dengue virus type 3 envelope glycoprotein.** *J Virol* 2005, **79(2)**:1223-1231.
29. Zhang S, Bovshik EI, Maillard R, Gromowski GD, Volk DE, Schein CH, Huang CY, Gorenstein DG, Lee JC, Barrett AD, *et al*: **Role of BC loop residues in structure, function and antigenicity of the West Nile virus envelope protein receptor-binding domain III.** *Virology* 2010, **403(1)**:85-91.
30. Kuno G, Chang GJ, Tsuchiya KR, Karabatsos N, Cropp CB: **Phylogeny of the genus Flavivirus.** *J Virol* 1998, **72(1)**:73-83.
31. Seligman SJ: **Constancy and diversity in the flavivirus fusion peptide.** *Virol J* 2008, **5**:27.
32. Gould EA, de Lamballerie X, Zanotto PMD, Holmes EC: **Origins, evolution, and vector/host coadaptations within the genus Flavivirus.** *Flaviviruses: Structure, Replication and Evolution* 2003, **59**:277-+.
33. Grard G, Moureau G, Charrel RN, Lemasson JJ, Gonzalez JP, Gallian P, Gritsun TS, Holmes EC, Gould EA, de Lamballerie X: **Genetic characterization of tick-borne flaviviruses: New insights into evolution, pathogenetic determinants and taxonomy.** *Virology* 2007, **361(1)**:80-92.
34. Crabtree MB, Sang RC, Stollar V, Dunster LM, Miller BR: **Genetic and phenotypic characterization of the newly described insect flavivirus, Kamiti River virus.** *Arch Virol* 2003, **148**:1095-1118.
35. Roehrig JT: **Antigenic structure of flavivirus proteins.** *Adv Virus Res* 2003, **59**:141-175.
36. Babu JP, Pattnaik P, Gupta N, Shrivastava A, Khan M, Rao PVL: **Immunogenicity of a recombinant envelope domain III protein of dengue virus type-4 with various adjuvants in mice.** *Vaccine* 2008, **26(36)**:4655-4663.
37. Chin JFL, Chu JJH, Ng ML: **The envelope glycoprotein domain III of dengue virus serotypes 1 and 2 inhibit virus entry.** *Microbes and Infection* 2007, **9(1)**:1-6.
38. Gromowski GD, Barrett ND, Barrett AD: **Characterization of dengue virus complex-specific neutralizing epitopes on envelope protein domain III of dengue 2 virus.** *J Virol* 2008, **82(17)**:8828-8837.
39. Guy B, Guirakhoo F, Barban V, Higgs S, Monath TP, Lang J: **Preclinical and clinical development of YFV 17D-based chimeric vaccines against dengue, West Nile and Japanese encephalitis viruses.** *Vaccine* 2010, **28(3)**:632-649.
40. Matsui K, Gromowski GD, Li L, Schuh AJ, Lee JC, Barrett AD: **Characterization of dengue complex-reactive epitopes on dengue 3 virus envelope protein domain III.** *Virology* 2009, **384**:16-20.
41. Murrell S, Wu S-C, Butler M: **Review of dengue virus and the development of a vaccine.** *Biotechnology Advances* 2011, **29(2)**:239-247.
42. Ramanathan MP, Kuo YC, Selling BH, Li Q, Sardesai NY, Kim JJ, Weiner DB: **Development of a novel DNA SynCon tetravalent dengue vaccine that elicits immune responses against four serotypes.** *Vaccine* 2009, **27(46)**:6444-6453.
43. Webster DP, Farrar J, Rowland-Jones S: **Progress towards a dengue vaccine.** *The Lancet Infectious Diseases* 2009, **9(11)**:678-687.
44. Rico-Hesse R, Harrison LM, Nisalak A, Vaughn DW, Kalayanarooj S, Green S, Rothman AL, Ennis FA: **Molecular evolution of dengue type 2 virus in Thailand.** *Am J Trop Med Hyg* 1998, **58(1)**:96-101.
45. Barker WC, Mazumder R, Vasudevan S, Sagripanti JL, Wu CH: **Sequence signatures in envelope protein may determine whether flaviviruses produce hemorrhagic or encephalitic syndromes.** *Virus Genes* 2009, **39(1)**:1-9.
46. Oliphant T, Nybakken GE, Engle M, Xu Q, Nelson CA, Sukupolvi-Petty S, Marri A, Lachmi BE, Olshevsky U, Fremont DH, *et al*: **Antibody recognition and neutralization determinants on domains I and II of West Nile virus envelope protein.** *Journal of Virology* 2006, **80(24)**:12149-12159.
47. Sukupolvi-Petty S, Austin SK, Purtha WE, Oliphant T, Nybakken GE, Schlesinger JJ, Roehrig JT, Gromowski GD, Barrett AD, Fremont DH, *et al*: **Type- and subcomplex-specific neutralizing antibodies against domain III of dengue virus type 2 envelope protein recognize adjacent epitopes.** *Journal of Virology* 2007, **81(23)**:12816-12826.
48. Danecek P, Lu W, Schein CH: **PCP consensus sequences of flaviviruses: correlating variance with vector competence and disease phenotype.** *J Mol Biol* 2010, **396(3)**:550-563.
49. Przybylski D, Rost B: **Powerful fusion: PSI-BLAST and consensus sequences.** *Bioinformatics* 2008, **24(18)**:1987-1993.
50. Davis CT, Ebel GD, Lanciotti RS, Brault AC, Guzman H, Siirin M, Lambert A, Parsons RE, Beasley DW, Novak RJ, *et al*: **Phylogenetic analysis of North American West Nile virus isolates, 2001-2004: evidence for the emergence of a dominant genotype.** *Virology* 2005, **342(2)**:252-265.
51. **Locally acquired Dengue–Key West, Florida, 2009-2010.** *MMWR Morb Mortal Wkly Rep* 2010, **59(19)**:577-581.
52. Gubler DJ: **Epidemic dengue/dengue hemorrhagic fever as a public health, social and economic problem in the 21st century.** *TRENDS Microbiol* 2002, **10(2)**:100-103.
53. Guzman MG, Kouri G: **Dengue and dengue hemorrhagic fever in the Americas: lessions and challenges.** *J Clin Virol* 2003, **27(1)**:1-13.
54. Rico-Hesse R: **Microevolution and virulence of dengue viruses.** *Adv Virus Res* 2003, **59**:315-341.
55. Streit J, Yang M, Cavanaugh J, Polgreen P: **Upward trend in dengue incidence among hospitalized patients, United States.** *Emerg Infect Dis* 2011, [serial on the Internet].
56. Bowen D, Lewis J, Lu W, Schein C: **Simplifying complex sequence information: a PCP-consensus protein binds antibodies against all four Dengue serotypes.** *Vaccine* 2012, (in press).
57. Schein CH, Oezguen N, van der Heden van Noort GJ, Filippov DV, Paul A, Kumar E, Braun W: **NMR solution structure of poliovirus uridylyated peptide linked to the genome (VPgpU).** *Peptides* 2010, **31(8)**:1441-1448.
58. Schein CH, Oezguen N, Volk DE, Garimella R, Paul A, Braun W: **NMR structure of the viral peptide linked to the genome (VPg) of poliovirus.** *Peptides* 2006, **27(7)**:1676-1684.
59. Schein CH, Volk DE, Oezguen N, Paul A: **Novel, structure-based mechanism for uridylylation of the genome-linked peptide (VPg) of picornaviruses.** *Proteins* 2006, **63(4)**:719-726.
60. Tuin AW, Siegal G, van der Marel GA, Overkleeft HS, Filippov DV: **Facile synthesis and application of uniformly 13C, 15N-labeled phosphotyrosine for ligand binding studies.** *Bioorg Med Chem Lett* 2006, **16(14)**:3806-3808.
61. van der Heden van Noort GJ, Overkleeft HS, van der Marel GA, Filippov DV: **Synthesis of nucleotidylated poliovirus VPg proteins.** *J Org Chem* 2010, **75(16)**:5733-5736.

doi:10.1186/1471-2105-13-S13-S9
**Cite this article as:** Schein *et al.*: Physicochemical property consensus sequences for functional analysis, design of multivalent antigens and targeted antivirals. *BMC Bioinformatics* 2012 **13**(Suppl 13):S9.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central