

Low dS/dN Does Not Correlate With High Variation of Amino Acid Sequences Along the gp120 Protein Structure

Zach Goldstein & Jordan Detamore
BIOL 368: Bioinformatics Laboratory
Department of Biology
Loyola Marymount University
November 15, 2016

Outline

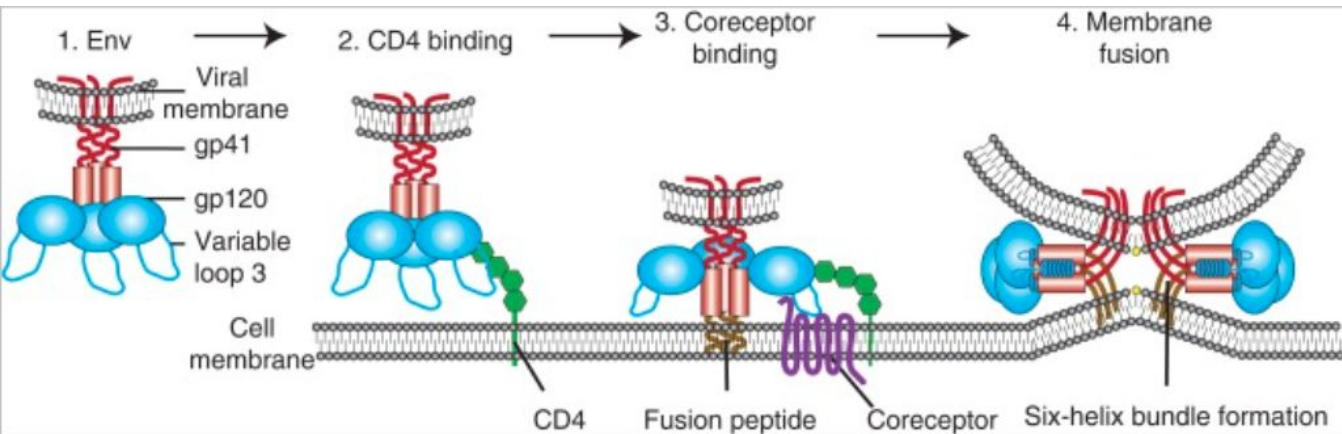
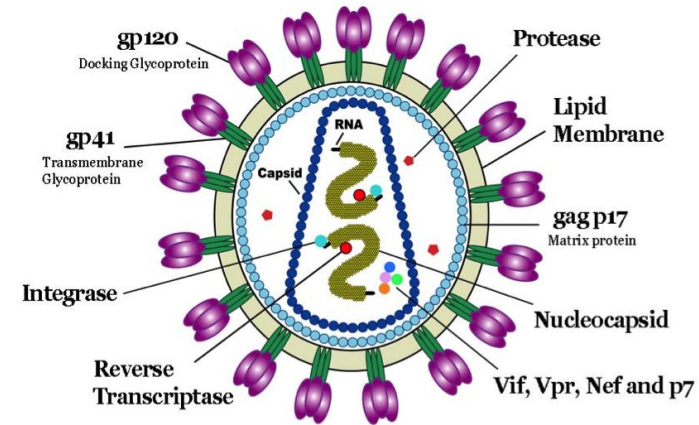
- **Low dS/dN ratios may correlate with high amino acid variations along the V3 loop of the gp120 protein**
- **Subjects selected from Markham et al. (1998) demonstrated high or low dS/dN values**
- **Amino acid sequence alignment using ClustalW and BoxShade showed conflicting results**
- **6 Non-consensus segments along gp120 are found along the V3 loop and on the exterior surface of the protein**
- **Further research would examine the impact of variations in these 6 non-consensus sites**

Outline

- **Low dS/dN ratios may correlate with high amino acid variations along the V3 loop of the gp120 protein**
- Subjects selected from Markham et al. (1998) demonstrated high or low dS/dN values
- Amino acid sequence alignment using ClustalW and BoxShade showed conflicting results
- 6 Non-consensus segments along gp120 are found along the V3 loop and on the exterior surface of the protein
- Further research would examine the impact of variations in these 6 non-consensus sites

HIV is a Retrovirus With Many Different Structures

- **gp120:** Facilitates HIV entry into host cell by binding to CD4
- **gp41:** Responsible for fusion of viral and target cell membranes
- **V3 Region:** Highly variable region of gp120 that facilitates coreceptor binding



http://people.mpi-inf.mpg.de/~fmueeller/thesis/thesis_mueller.pdf

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3405824/>

Non-Synonymous Mutation Rates May Demonstrate Significant Changes in Amino Acid Sequence

- Non synonymous mutations (dN) are nucleotide changes that lead to the coding of a different amino acid.
- Synonymous mutations (dS) are nucleotide changes that do not lead to the coding of a different amino acid.
- These ratios were formed for each consensus strain for all subjects.
- The smaller the ratio, the larger amount of non synonymous mutations, the more potential for differences in amino acid sequences

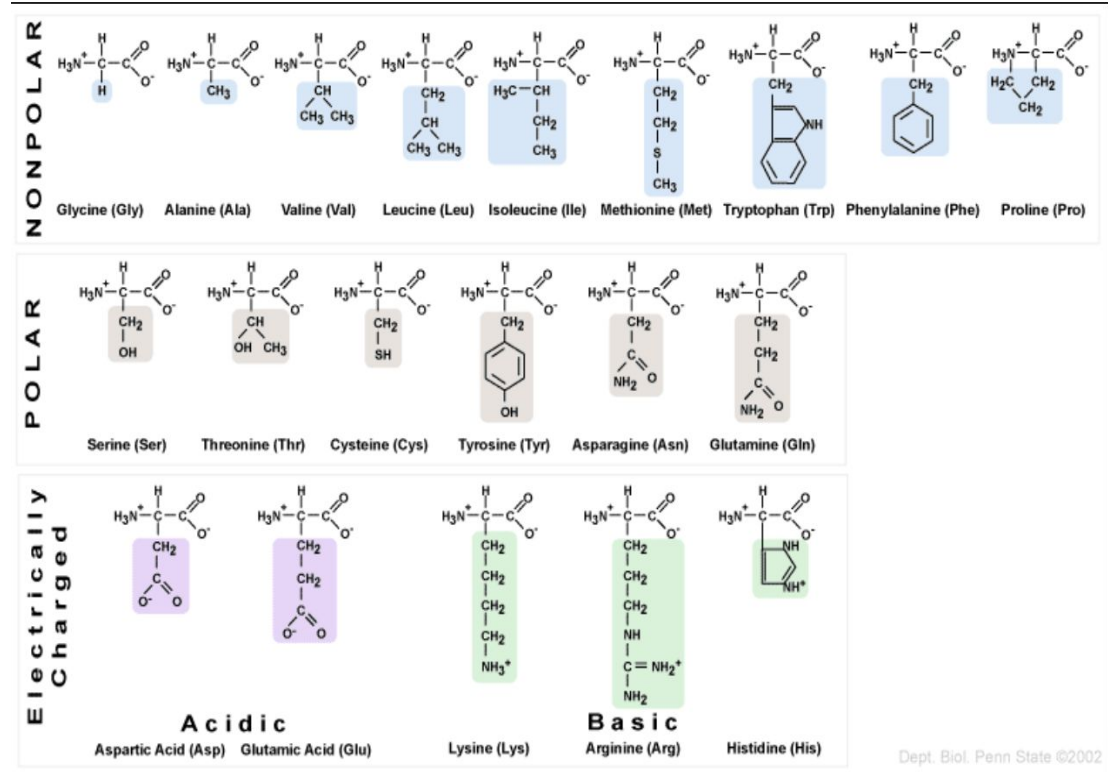
20 Standard Amino Acids Vary in Structure and Properties

4 Groups of Amino Acids

- Acidic
- Basic
- Uncharged Polar
- Nonpolar

Position affects function

Size affects function



Do Mutation Rates Affect Variation in Amino Acid Sequences?

Questions:

Does a low dS/dN ratio value result in high amino acid variation?

Do subjects with low dS/dN ratios vary from each other?

Do the sites of variation have nonconservative mutations?

Hypotheses:

Low dS/dN ratio values will result in high amino acid variation.

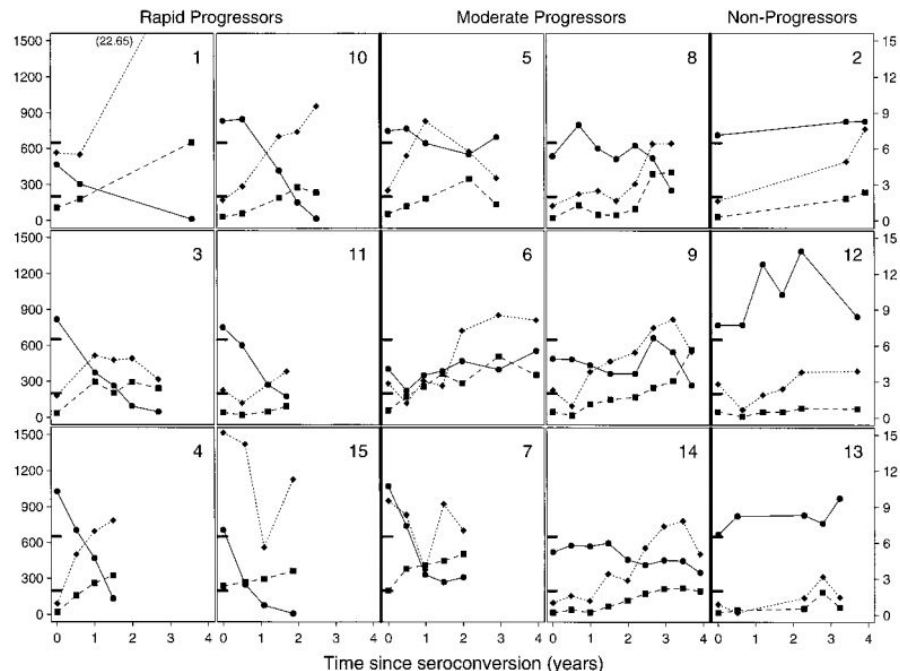
Subjects with low dS/dN ratios will vary greatly from each other.

The sites of high variation will result in nonconservative mutations.

Outline

- Low dS/dN ratios may correlate with high amino acid variations along the V3 loop of the gp120 protein
- **Subjects selected from Markham et al. (1998) demonstrated high or low dS/dN values**
- Amino acid sequence alignment using ClustalW and BoxShade showed conflicting results
- 6 Non-consensus segments along gp120 are found along the V3 loop and on the exterior surface of the protein
- Further research would examine the impact of variations in these 6 non-consensus sites

Subjects Selected Based on Differences in dS/dN Ratios



- Four subjects with highest dS/dN ratio and four subjects with lowest dS/dN ratio were chosen
- Three random clones were selected from the first and last visit
- Subjects had different numbers of visits spread over the four years
 - Unable to choose subjects from same visit date

New Groups Formed Based on Highest and Lowest dS/dN Ratios

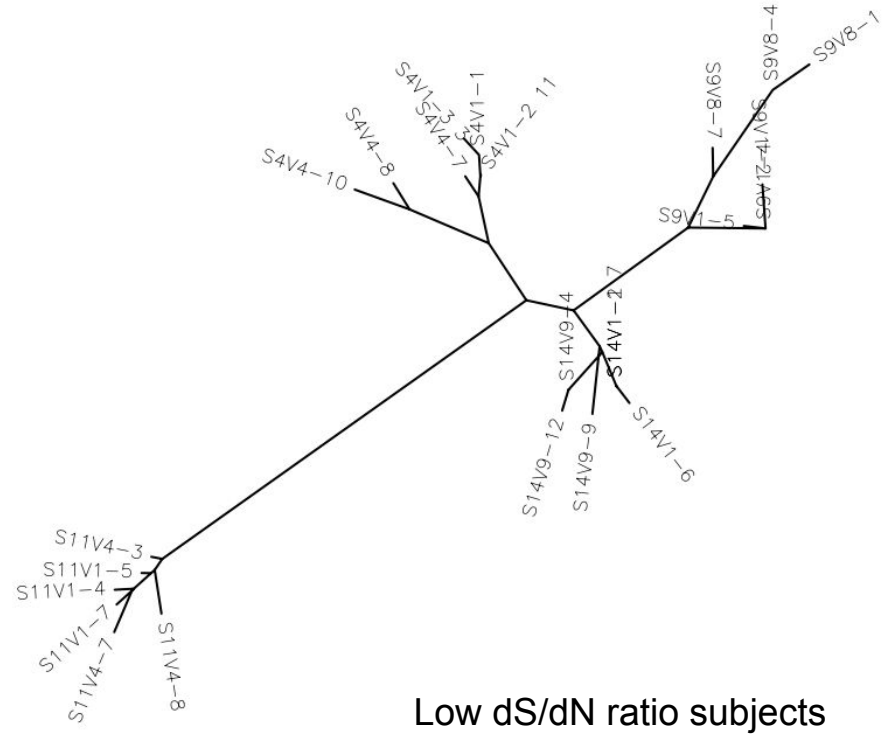
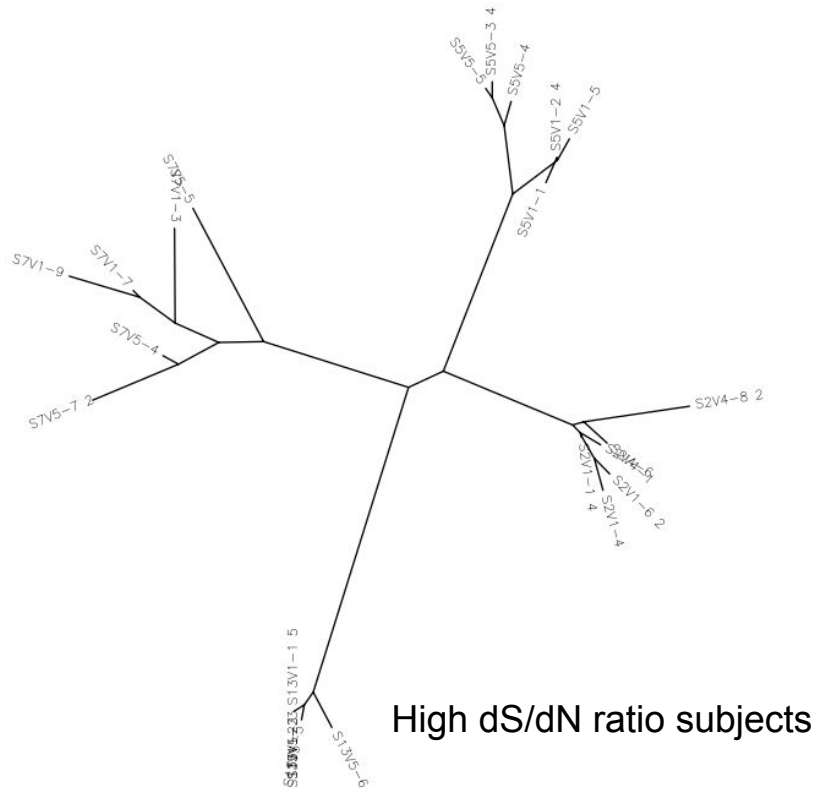
Rapid Progressor (Subject #)	Moderate Progressor (Subject #)	Non-progressor (Subject #)
4	7	2
10	8	12
11	14	13
15	5	
3	9	
1	6	

Subjects demonstrating highest dS/dN ratios	Subjects demonstrating lowest dS/dN ratios
7	4
5	9
2	11
13	14

Outline

- Low dS/dN ratios may correlate with high amino acid variations along the V3 loop of the gp120 protein
- Subjects selected from Markham et al. (1998) demonstrated high or low dS/dN values
- **Amino acid sequence alignment using ClustalW and BoxShade showed conflicting results**
- 6 Non-consensus segments along gp120 are found along the V3 loop and on the exterior surface of the protein
- Further research would examine the impact of variations in these 6 non-consensus sites

Phylogenetic Trees Show No Clear Difference in Variation Between Low dS/dN Subjects and High dS/dN Subjects



BoxShade and ClustalW Were Used to Analyze Amino Acid Sequences

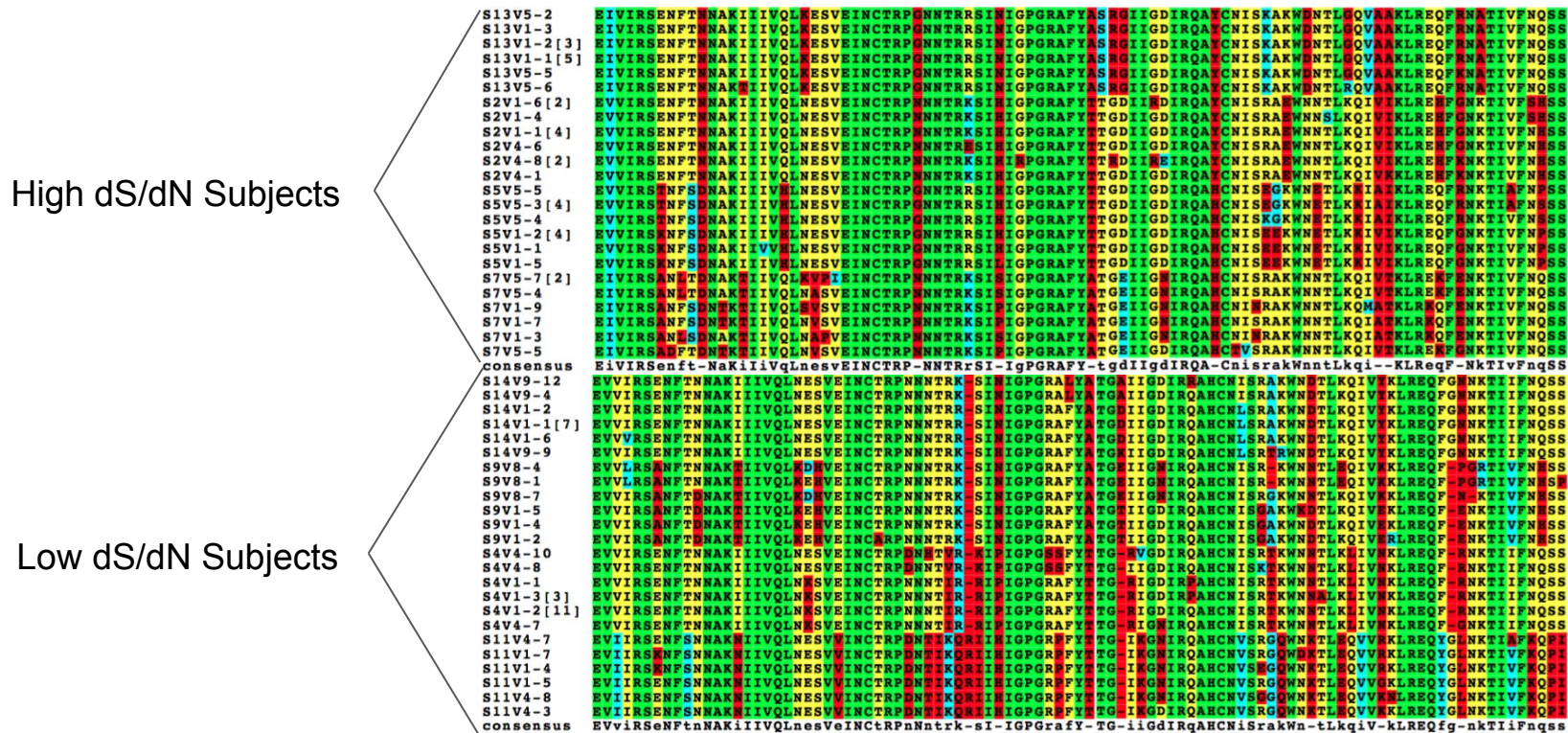
Boxshade default similarities	Individual Similarities: D: E F: YW G: A I: LVM L: VMI M: ILV N: Q R: K T: S V: MIL W: FY Y: WF Groups: FYW IVLM RK DE GA TS NQ
	Individual Similarities: A: GS D: EN E: DQ F: YW G: A H: RK I: LVM K: HR L: VMI N: QD Q: NE R: KH S: TA T: S V: MIL W: FY Y: WF Groups: ACST NT EQ NQ ED KRQH GN NH FY IVLM

Completely Conserved Residues
Background Color: Green
Identical Residues
Background Color: Yellow
Similar Residues
Background Color: Cyan
Different Residues
Background Color: Red

Low Ratio Subjects Had More Non-Consensus Sites than High Ratio Subjects

	High dS/dN Ratio Subjects	Low dS/dN Ratio Subjects	All Subjects
Fully Conserved Residue	49	43	30
Conservation of Strong Group	22	20	23
Conservation of Weak Group	10	6	13
No Consensus	14	24	31

BoxShade Highlights a Higher Frequency of Non-Consensus Sites in Subjects From the High dS/dN Group



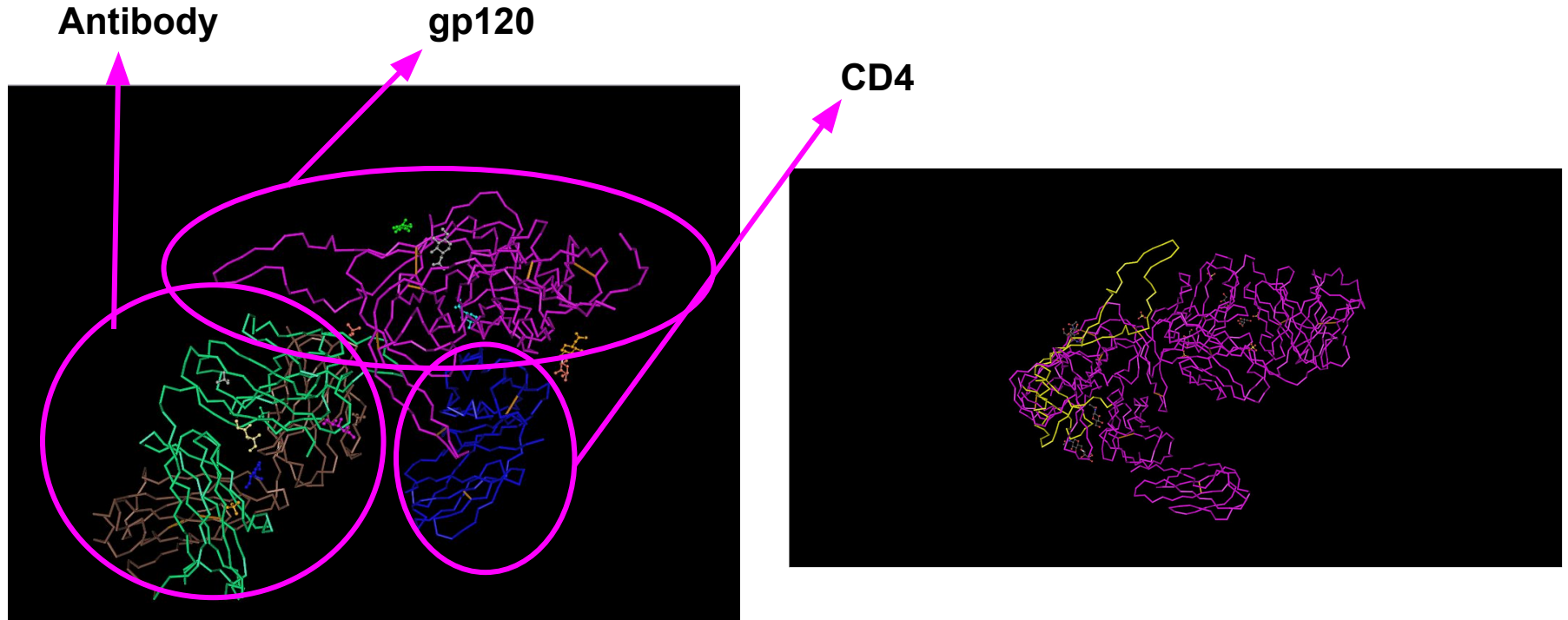
Outline

- Low dS/dN ratios may correlate with high amino acid variations along the V3 loop of the gp120 protein
- Subjects selected from Markham et al. (1998) demonstrated high or low dS/dN values
- Amino acid sequence alignment using ClustalW and BoxShade showed conflicting results
- **6 Non-consensus segments along gp120 are found along the V3 loop and on the exterior surface of the protein**
- Further research would examine the impact of variations in these 6 non-consensus sites

Comparison of all Subjects Show 6 Sites With no Amino Acid Consensus

S13V5-2 EVVIRSENFNTNNAKIIIVQLNESVEINCTRPNNNTRRSINIGPGRAFY TGGIIGDTRQAACNISAKKWNNTLQVAAKIKLEQFNKTTIVFNQSSS
 S13V1-3 EVVIRSENFNTNNAKIIIVQLNESVEINCTRPNNNTRRSINIGPGRAFY TGGIIGDTRQAACNISAKKWNNTLQVAAKIKLEQFNKTTIVFNQSSS
 S13V1-2[3] EVVIRSENFNTNNAKIIIVQLNESVEINCTRPNNNTRRSINIGPGRAFY TGGIIGDTRQAACNISAKKWNNTLQVAAKIKLEQFNKTTIVFNQSSS
 S13V1-1[5] EVVIRSENFNTNNAKIIIVQLNESVEINCTRPNNNTRRSINIGPGRAFY TGGIIGDTRQAACNISAKKWNNTLQVAAKIKLEQFNKTTIVFNQSSS
 S13V5-5 EVVIRSENFNTNNAKIIIVQLNESVEINCTRPNNNTRRSINIGPGRAFY TGGIIGDTRQAACNISAKKWNNTLQVAAKIKLEQFNKTTIVFNQSSS
 S13V5-6 EVVIRSENFNTNNAKIIIVQLNESVEINCTRPNNNTRRSINIGPGRAFY TGGIIGDTRQAACNISAKKWNNTLQVAAKIKLEQFNKTTIVFNQSSS
 S9V8-4 EVVILRSANFTNNAKTIIVQLDVEINCTRPNNNTRKSINIGPGRAFY TGEIIGNTRQAHCNISRAKWNNTLKQIVKIKLEQFNGRTIVFNSSS
 S9V8-1 EVVILRSANFTNNAKTIIVQLDVEINCTRPNNNTRKSINIGPGRAFY TGEIIGNTRQAHCNISRAKWNNTLKQIVKIKLEQFNGRTIVFNSSS
 S9V8-7 EVVIRSANFTDNAKTIIVQLDVEINCTRPNNNTRKSINIGPGRAFY TGEIIGNTRQAHCNISRAKWNNTLKQIVKIKLEQFNGRTIVFNSSS
 S9V1-4 EVVIRSANFTDNAKTIIVQLNESVEINCTRPNNNTRKSINIGPGRAFY TGGIIGDTRQAACNISAKKWNNTLKQIVKIKLEQFNKTTIVFNSSS
 S9V1-2 EVVIRSANFTDNAKTIIVQLNESVEINCTRPNNNTRKSINIGPGRAFY TGGIIGDTRQAACNISAKKWNNTLKQIVKIKLEQFNGRTIVFNSSS
 S9V1-5 EVVIRSANFTDNAKTIIVQLNESVEINCTRPNNNTRKSINIGPGRAFY TGGIIGDTRQAACNISAKKWNNTLKQIVKIKLEQFNGRTIVFNSSS
 S7V5-7[2] EVVIRSANFTDNAKTIIVQLVLEINCTRPNNNTRKSISIGPGRAFY TGEIIGNTRQAHCNISRAKWNNTLKQIVKIKLEQFNKTTIVFNQSSS
 S7V5-4 EVVIRSANFTDNAKTIIVQLNASVEINCTRPNNNTRKSISIGPGRAFY TGEIIGNTRQAHCNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S7V1-9 EVVIRSANFSDNAKTIIVQLNVSVVEINCTRPNNNTRKSISIGPGRAFY TGEIIGNTRQAHCNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S7V1-7 EVVIRSANFSDNAKTIIVQLNVSVVEINCTRPNNNTRKSISIGPGRAFY TGEIIGNTRQAHCNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S7V1-3 EVVIRSANFSDNAKTIIVQLNVSVVEINCTRPNNNTRKSISIGPGRAFY TGEIIGNTRQAHCNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S7V5-5 EVVIRSANFTDNAKTIIVQLNVSVVEINCTRPNNNTRKSISIGPGRAFY TGEIIGNTRQAHCNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S14V9-12 EVVIRSENFNTNNAKIIIVQLNESVEINCTRPNNNTRRSINIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S14V9-4 EVVIRSENFNTNNAKIIIVQLNESVEINCTRPNNNTRRSINIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S14V1-2 EVVIRSENFNTNNAKIIIVQLNESVEINCTRPNNNTRRSINIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S14V1-1[7] EVVIRSENFNTNNAKIIIVQLNESVEINCTRPNNNTRRSINIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S14V1-6 EVVIRSENFNTNNAKIIIVQLNESVEINCTRPNNNTRRSINIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S14V9-9 EVVIRSENFNTNNAKIIIVQLNESVEINCTRPNNNTRRSINIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S2V1-6[2] EVVIRSENFNTNNAKIIIVQLNESVEINCTRPNNNTRKSINIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S2V1-4 EVVIRSENFNTNNAKIIIVQLNESVEINCTRPNNNTRKSINIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S2V4-8[2] EVVIRSENFNTNNAKIIIVQLNESVEINCTRPNNNTRKSINIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S2V1-1[4] EVVIRSENFNTNNAKIIIVQLNESVEINCTRPNNNTRKSINIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S2V4-6 EVVIRSENFNTNNAKIIIVQLNESVEINCTRPNNNTRKSINIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S2V4-1 EVVIRSENFNTNNAKIIIVQLNESVEINCTRPNNNTRKSINIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S4V4-10 EVVIRSENFNTNNAKIIIVQLNESVEINCTRPNNNTRRIPIIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S4V4-8 EVVIRSENFNTNNAKIIIVQLNESVEINCTRPNNNTRRIPIIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S4V1-1 EVVIRSENFNTNNAKIIIVQLNESVEINCTRPNNNTRRIPIIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S4V1-3[3] EVVIRSENFNTNNAKIIIVQLNESVEINCTRPNNNTRRIPIIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S4V1-2[11] EVVIRSENFNTNNAKIIIVQLNESVEINCTRPNNNTRRIPIIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S4V4-7 EVVIRSENFNTNNAKIIIVQLNESVEINCTRPNNNTRRIPIIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S5V5-5 EVVIRSANFSDNAKIIIVQLNESVEINCTRPNNNTRRSINIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S5V5-3[4] EVVIRSANFSDNAKIIIVQLNESVEINCTRPNNNTRRSINIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S5V5-4 EVVIRSANFSDNAKIIIVQLNESVEINCTRPNNNTRRSINIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S5V1-2[4] EVVIRSANFSDNAKIIIVQLNESVEINCTRPNNNTRRSINIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S5V1-1 EVVIRSANFSDNAKIIIVQLNESVEINCTRPNNNTRRSINIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S5V1-5 EVVIRSANFSDNAKIIIVQLNESVEINCTRPNNNTRRSINIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S11V4-7 EVVIRSENFNTNNAKIIIVQLNESVEINCTRPNNNTRKRIPIIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S11V1-7 EVVIRSANFSDNAKIIIVQLNESVEINCTRPNNNTRKRIPIIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S11V1-4 EVVIRSANFSDNAKIIIVQLNESVEINCTRPNNNTRKRIPIIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S11V1-5 EVVIRSANFSDNAKIIIVQLNESVEINCTRPNNNTRKRIPIIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S11V4-8 EVVIRSENFNTNNAKIIIVQLNESVEINCTRPNNNTRKRIPIIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 S11V4-3 EVVIRSENFNTNNAKIIIVQLNESVEINCTRPNNNTRKRIPIIGPGRAFY TGGIIGDTRQAACNISRAKWNNTLKQIVKIKLEQFNGRTIVFNQSSS
 consensus EvvIrSenfntnNaKiiIvQLnesveINcTrPnNtr-rS-i-IGPGRAFY-tgeiigDTRQAACNisrAKWNntLKQIV-kLEQF-n-NTIvFNqssS

The V3 Region Sticks Out of gp120 Facing the Antibody

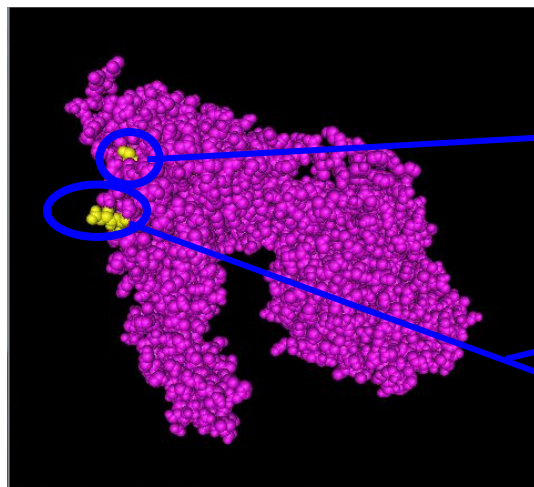


Huang et al. (2005)

Three Sites Along the V3 Region Show Variability

Along the Outside of the Protein

Key
A: Acidic
B: Basic
P: Polar
NP: Nonpolar



Lysine (B) → Glycine (NP), Arginine (B),
 Isoleucine (NP), Asparagine (P), Tyrosine (P),
 Glutamic Acid (A), Threonine (P), Alanine (NP)

Asparagine (P) → Glycine (NP), Arginine (B),
 Lysine (B), Glutamic Acid (A), Proline (NP)

Lysine (B) → Leucine (NP), Asparagine (P),
 Insertion

Cn3D macromolecular structure viewer

234C_G	vqcthgirpvytqlllngslaeeevirsdnfnnaktiivqlkesveinctrpnqnrksihigpgrafyttgeiigdirqahcnisrakwndtlkqivi
234C_C	nsdthllqgqsltltlesppgsspsvqcrsprgkniqggkltlsvsqtelqdsqgtwtctvlqnqkkvefkidivvlafqk
234C_I	tkleikrtvaapsvfifppsdeqlksqgtasvvcflnnfyprcakvqkwvdnalqsgnsqesvteqdsrdstyslgstltliskadyekkhvyacevthqglsspvtksfnrgec
234C_H	pdwedgdsydgsgrgffdfwgqgtlvtvssastkgpsvfpapsskstsggtaalgclvkdyfpepvtvswngaltsgvhtfpaviqssglyslssvvtvpsssgtqtyicnvnhkpsntkvdkkvepksc

Three Sites Along the immunogenetic tip of the V3 Region Show High Variability

Key
A: Acidic
B: Basic
P: Polar
NP: Nonpolar

Lysine (B) → Glutamine (P) or insertion

Isoleucine (NP) → Asparagine (P), Proline (NP),
Serine (P), Histidine (B), Leucine (NP)

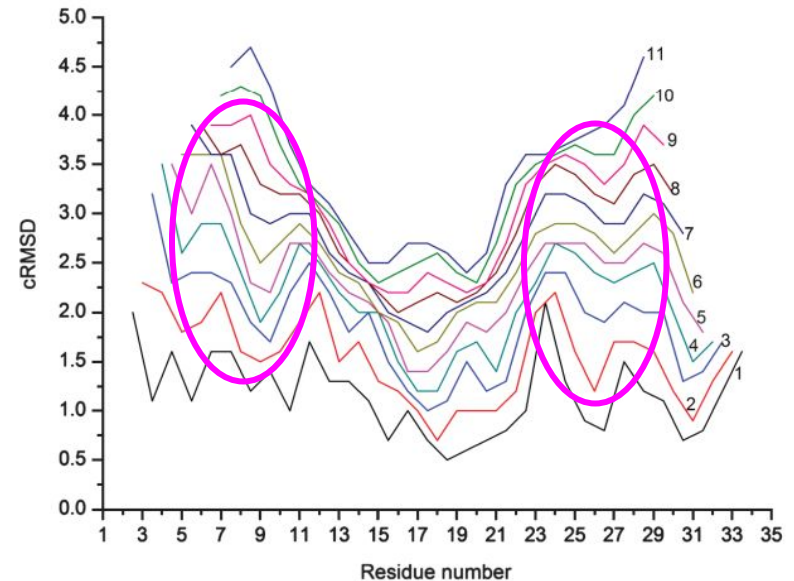
Threonine (P) → Alanine (NP)

Cn3D macromolecular structure viewer

```
2B4C_G vqcthgirpvvtqllngslaeeevirsdnftnnaktiivqlkesveinctrpnqntfshihgpggrafytigeiigdirqahcnisrakwndtlkqiviklrqfenktivfnhssggdpeivmhsfncggeffycnsaqlfr
2B4C_C nsdthl1qgqsltltlesppgsspsvqcrsprgkniqggktilsvsqlelqdsqgtwtctvlqnqkkvefkidivv1afqk
2B4C_I tkleikrtvaapsvfifppsdeqlksgtasvvc1lnnfypreakvqkwvdnalqsgnsqesvteqdsrdstyslgsstl1skadyekkhkvyacevthqglsspvtksfnrgec
2B4C_H pdwedgdsydgsgrgffdfwgqgtlvtvssastkgpsvfp1apsakstsggtaalgclvkdyfpepvtvswnsгалtsgvhtfpaviqssglyslsavvtvpssslgtqtyicnvnhkpsntkvdkkvepksc
```

Non-Consensus Regions of gp120 Match Variable Segments Studied by Andrianov & Anishchenko (2012)

- A 35-residue-long segment of the V3 loop shows high structure variability
- 3 segments along the same V3 region found to be highly variable
- Conservation of amino acid composition found at “pinch” sites are targets for anti-AIDS medication



Low dS/dN Ratios Demonstrate no Clear Relationship to Higher Rates of Non-Consensus Segments Along the V3 Loop

- The sites of high variation result in nonconservative mutations, causing changes in protein structure and function
- Interactions and functions of protein may be affected by variations in primary sequence
- Non-Consensus segments found along V3 loop and on the outside region of gp120 using BoxShade were not consistent with ClustalW

Outline

- Low dS/dN ratios may correlate with high amino acid variations along the V3 loop of the gp120 protein
- Subjects selected from Markham et al. (1998) demonstrated high or low dS/dN values
- Amino acid sequence alignment using ClustalW and BoxShade showed conflicting results
- 6 Non-consensus segments along gp120 are found along the V3 loop and on the exterior surface of the protein
- **Further research would examine the impact of variations in these 6 non-consensus sites**

Functions of Amino Acid Variations Must be Studied to Observe the Effect of Non-Consensus Sites

- Low dS/dN ratio subjects show trend towards significant rates of mutations causing variable amino acid structure using BoxShade
- Specific functions of amino acid substitutions along the V3 loop should be studied
- V3 region should be targeted for anti-AIDS medications due to rigidity in structure, studying variations in structure may prove helpful
- Data can be used to identify HIV-1 immune system evasion

Summary

- **Low dS/dN ratios may correlate with high amino acid variations along the V3 loop of the gp120 protein**
- **Subjects selected from Markham et al. (1998) demonstrated high or low dS/dN values**
- **Amino acid sequence alignment using ClustalW and BoxShade showed conflicting results**
- **6 Non-consensus segments along gp120 are found along the V3 loop and on the exterior surface of the protein**
- **Further research would examine the impact of variations in these 6 non-consensus sites**

Acknowledgements

Dr. Kam Dahlquist

Department of Biology at Loyola Marymount University

Bioinformatics Lab Fall 2016



References

Andrianov, A. M., & Anishchenko, I. V. (2009). Computational model of the HIV-1 subtype A V3 loop: Study on the conformational mobility for structure-based anti-AIDS drug design. *Journal of Biomolecular Structure and Dynamics*, 27(2), 179-193. DOI: 10.1080/07391102.2009.10507308

Huang CC et al. *Science* (2005) 310 p.1025. Structure of a v3-containing hiv-1 gp120 core.

Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J., & Hendrickson, W. A. (1998). Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature*, 393(6686), 648-659. DOI: 10.1038/31405

Markham, R.B., Wang, W.C., Weisstein, A.E., Wang, Z., Munoz, A., Templeton, A., Margolick, J., Vlahov, D., Quinn, T., Farzadegan, H., & Yu, X.F. (1998). Patterns of HIV-1 evolution in individuals with differing rates of CD4 T cell decline. *Proc Natl Acad Sci U S A*. 95, 12568-12573. doi: 10.1073/pnas.95.21.12568

Cn3D macromolecular structure viewer

Biology Workbench