A Preliminary Study on Human Immune Response by Changing the Potential Energy and Hydrophobicity of HIV Membrane Protein

Liping Zhao, Qiliang Shu

Jingdezhen Ceramic Institute Information Engineering Department, China.

Abstract

By analyzing the potential energy and hydrophobicity of HIV membrane proteins, a new method is proposed to effectively induce human immune response by changing the cracks between the gp120 protein molecules and the outer domain so that the antigen site is fully exposed.

Keywords

HIV, potential hydrophobic, gp120.

1. Introduction

The structure of the pre-HIV virus genome consists of two linear RNAs, both of which form dimmers through hydrogen bonds. The total length of the provirus gene is about 9.7 KB. Both sides of the genome are non-coding regions, that is, long-terminal repeat sequences(LTRs). The central region between the non-coding regions on both sides is the protein-coding region. It consists of three structural genes, gag, pol, and env. Consisting, it is common for retroviruses. Encoding viral core proteins(gag), polymerase (Pol), and outer membrane proteins(env), respectively. The six regulatory bases are vpr, rev, vif, tat, vpu/vpx, and NEF, and each gene can encode one or more proteins. Regulatory genes often overlap in structural genes. The genetic structure of HIV-1 and HIV-2 is basically the same, except that VPU is unique to HIV-1 and VPX is unique to HIV-2.

CD4 is the most important target cell for HIV, including CD4 + T cells and mononuclear macrophages. HIV can adhere to CD4 molecules on the surface of the cell and then induce membrane fusion into the target cell through conformation changes. In CD4 lymphocyte intracellular, RNA is reversed to provirus DNA, which is then integrated into the host cell genome DNA and then copied with the cell.

A large amount of biochemical and structural biological evidence shows that the infection of HIV-1 virus to cells is initiated by gp120 molecules: First, gp120 exposes the receptor binding site through interaction with receptor CD4; Second, the further interaction between gp120 and co-receptor causes the gp120 to undergo conformational rearrangement, causing the HIV-1 Tran membrane protein gp41 to insert into the cell membrane, causing the viral membrane to fuse with the target cell membrane and the virus to enter the cell. This process involves multiple rearrangement of gp120 molecules and the structural dynamics of gp120 molecules play a very important role in the leaching process.

At present, vaccines are considered to be the most effective way to control the spread of HIV. Gag protein is one of the main structural proteins of HIV. Because the amino acid sequence of the AG protein is relatively conservative and the antigen variation is less, the use of the AG protein as an AIDS vaccine may be able to overcome the defect that the enV protein can not effectively resist the attack of alien mutant virus strains. At the same time, studies have shown that the AG protein contains many antigen-determining clusters of dominant antigens that can cause the body to produce specific humeral immunity(including neutralizing antibodies) and specific CTL reactions. In addition, another important feature of the AG protein is its self-assembly function, which can self-assemble virus-like particles.

This is very necessary for the construction of macromolecular granulated antigens, so gag protein has become a new hotspot in the research of HIV vaccine, especially granule vaccine. Bactrian Bacillus is a widely used gene expression system in recent years. It has many advantages such as high expression rate, genetic stability, product secretion, and mature fermentation process.

In HIV/AIDS patients in China, CD4BS, CD4i, and 2G12 neutralizing antibody conserved amino acids have mutated during the HIV infection period. The disease progresses to the AIDS stage, and the chance of further mutations is lower.; In the long-term non-progress of neutralizing antibody site, only 2G12 neutralizing antibody site amino acids were found to have mutations. In the gp120 C2-C3 region mutation site distribution, the ratio of 2G12 site was significantly higher than that of CD4BS and CD4i site. Neutral antibody site amino acid mutations are dominated by single point mutations; The variation degree of each amino acid site in different types of neutralization antibody conserved site was different.

2. Research Technique

First, changes gp120 protein intermolecular and external structural cracks, so that the antigen site is fully exposed, effectively causing the human immune response.

Change the cracks between internal and external domains by changing the overall polarity of the areas on both sides of the crack.

1, the polar properties of the amino acid side chain are in most cases determined by the second nucleoside of the genetic code. The second nucleoside is the Non-Polar side chain of pyrimidine, and the second side chain is polar when the second nucleoside is purine, but if the first nucleoside is U or A, if the second nucleoside is C, the side chain is no longer Non-Polar and polar. The second nucleoside is U that has absolute specificity to the non-polarity of the side chain. Only the Try set encoding(UGG) does not meet this principle.

2, the genetic code with the following characteristics determines that the amino acid side chain is polar non-charged. The first nucleoside must be U or A(note that U and A have complementary structural conditions). Under this premise, the second nucleoside is C, and the third nucleoside has no specificity; The second nucleoside is purine and the third nucleoside is pyrimidine. Only the two groups of codes CAA and CAG that do not meet this principle are Gin.

3, charged side chain amino acid coding is very concentrated, its feature is that the second nucleoside is purine and the first nucleoside can not be U. Under this premise, if the first nucleoside is C(except Gin's two groups of codes CAA, CAG) or the first nucleoside is A, the third nucleoside is also purine and the positive side chain amino acid is encoded; The first and second nucleosides determine the negative charge of the amino acid side chain respectively.

(II) The role of different forces in the folding of proteins and in maintaining the spatial structure of proteins

Including non-covalent bonds such as hydrophobic, electrostatic, hydrogen bonds, and Van der Waals forces, hydrophobic interactions are important forces that constitute remote interactions between residues. In addition, the formation of disulfide bonds between cysteine also played a decisive role in determining the protein conformation.

Contact map: A polypeptide chain of length N can be expressed as a matrix S of NxN. The elements in this matrix can be positioned as the following formula:

$$S_{ij} = \begin{cases} 1 & \text{If the residual I is in contact with the residual j} \\ 0 & \text{Otherwise} \end{cases} \quad (1)$$

The contact between the two residues is defined in the form of "". If the distance between the atoms of the two residues is less than one threshold, then the two residues are in contact. If the position of the atoms of the residue I and residue J is expressed as a sum, respectively, there are the following definitions:

$$S_{ij} = \begin{cases} 1 & if \left| r_i - r_j \right| \le d_c, \left| i - j \right| \ge 8 \\ 0 & otherwise \end{cases} \quad (2)$$

This distance is based on the distance of the residue pair on this polypeptide chain and a given threshold value(measured as 6-8A), and when the two residues differ by 8 residues in the sequence, the two residues are In contact.

The residue Ile appears more often in the residues of the remote contact pair, and the other residues are cys, Val, Tyr, Trp, Phe and Leu. And hydrophobic residues mainly affect the remote interaction between residues.

The residues Val, Leu, Ala, Gly and Ile will be more easily accessible to other residues. The residues Trp rarely form remote contact with other residues, and the other residues are Cys, His, Get, and Gn in turn.

For a given protein sequence of any given structure, we can divide the possible amino acid residue pairs it contains into two categories. One is those with contact residues defined by the above formula, defined as positive samples; The other type is a pair of residues that do not have contact and is recorded as a negative sample. For each predicted residue pair, code using its conformational features(such as residual spatial distance, secondary structure, amino acid composition, sequence conservation, and related mutation analysis, etc.) and its physical and chemical properties, and then use such as neural networks, Support vector machines and other classifiers to study a given sample set, and finally use a good classifier to predict the protein residues of unknown structures and obtain our protein contact map results.

3. Research Objectives

HIV causes AIDS by binding, entering, and eventually leading to the death of T-assisted cells. T helper cells are immune cells that are necessary to fight ordinary bacteria and other pathogens. Because HIV greatly reduces the number of T-assisted cells, ordinary pathogens can also cause death. An effective HIV vaccine will induce the body to produce corresponding antibodies(specific immune system molecules), immunoglobulin, before the body is attacked by the virus. These antibodies will circulate in the blood, track and kill the HIV virus.

However, most of the antibodies produced by the body against HIV are ineffective. Because the surface of the HIV virus is covered with sugar molecules, the antibodies can not sneak in, and then block the use of the virus, bind with it, and eventually infect a variety of proteins in the cell. In fact, the situation is more complicated because HIV is constantly mutating, so any vaccine-induced antibody must have the ability to detect and destroy multiple HIV strains that actually exist. Analyze gp120 glycoprotein and find an effective method to inhibit further HIV infection.

References

[1] Jianhua Jia, Liuxia Zhang, Zi, Liu, et.al. pSumo-CD: Predicting sumoylation sites in proteins with covariance discriminant algorithm by incorporating sequence-coupled effects into general PseAAC. Bioinformatics, Accepted.doi:10.1093/ bioinformatics/ btw387. (SCI indexed, IF:5. 766).

[2] Xuan Xiao, Hong-Liang Zou, Wei-Zhong Lin . iMem-Seq: A Multi-label Learning Classifier for Predicting Membrane Proteins Types· Journal of Membrane Biology Mar 2015.

[3] Xuan Xiao,Jian-Liang Min, Wei-Zhong Lin. IDrug-Target: Predicting the interactions between drug compounds and target proteins in cellular networking via benchmark dataset optimization approach Journal of biomolecular Structure & Dynamics Dec 2014.

[4] Wang-Ren Qiu, Xuan Xiao , Wei-Zhong Lin. IUbiq-Lys: Prediction of lysine ubiquitination sites in proteins by extracting sequence evolution information via a gray system model Journal of biomolecular Structure & Dynamics,Sep 2014.

[5] Jianhua Jia， iCar-PseCp: identify carbonylation sites in proteins by Monto Carlo sampling and incorporating sequence coupled effects into general PseAAC，Oncotarget .May 2016.
[6] Jianhua Jia · pSuc-Lys: Predict lysine succinylation sites in proteins with PseAAC and ensemble random forest approach, Journal of Theoretical Biology , Jan 2016.
[7] Liping Zhao, Drug Target Prediction based on Human HIV-1 Protein-Protein Network, journal of theoretical and applied information technology, 2013.4.
[8] Jianhua Jia, Zi Liu, Xuan Xiao.et. al. pSuc-Lys: Predict lysine succinylation sites in proteins with PseAAC and ensemble random forest approach. Journal of Theoretical Biology.2016,394(4): 223-230.doi: 10.1016/j.jtbi.2016.01.020. (SCI:000379888800020, IF:2.049).
[9] Jianhua Jia, Zi Liu, Xuan Xiao. et. al. iPPBS-Opt: A sequence-based ensemble classifier foridentifying protein-protein binding sites by optimizing imbalanced training datasets. Molecules.2016, 21 (1), 95. doi:10.3390/molecules21010095. (SCI: 000369486800019, IF:2.465).
[10] Jianhua Jia, Zi Liu, Xuan Xiao. et.al. iSuc-PseOpt: Identifying lysine succinylation sites in proteins by incorporating sequence-coupling effects into pseudo components and optimizingim balanced training dataset. Analytical Biochemistry.2016, 497(3): 48-56. doi:10. 1016/j.ab. 2015.12.009. (SCI indexed:000370909900008, IF:2.243).
[11] Jianhua Jia, Zi Liu, Xuan Xiao. et.al. iPPI-Esml: An ensemble classifier for identifying the interactions of proteins by incorporating their physicochemical propertiers and wavelet transform into PseAAC. Journal of Theoretical Biology. 2015, 21(337):47-56. (SCI:000355241600005. IF:2.049).
[12] Jianhua Jia. Zi Liu, Xiang Chen. et.al. Prediction of Protein-Protein Interactions using Chaos Game Representation and wavelet transform via random forests. Genetics and Molecular Research. 2015,14(4):11791-11805.doi: 10.4238/2015.October.2.13. (SCI indexed： 000365922800013. IF: 0.764).
[13] Jianhua Jia, Xuan Xiao, Bingxiang Liu. Prediction of Protein-Protein Interactions with Physicochemical Descriptors and Wavelet Transform via Random Forests. Accepted by Journal of Laboratory Automation (JALA). 2016, 21(3): 368-377. (SCI: 000377095200003. IF:1.297).
[14] Jianhua Jia, Zi Liu, Xuan Xiao. et.al. Identification of protein-protein binding sites by incorporating the physicochemical properties and stationary wavelet transforms into pseudo amino acid composition. Accepted by Journal of Biomolecular Structure & Dynamics. 2016,34(9): 1946-1961.doi: 10.1080/07391102.2015.1095116. (SCI indexed. IF:2.300).
[15] Jianhua Jia, Xuan Xiao, Bingxiang Liu, et.al. Bagging-based Spectral Clustering Ensemble Selection. Pattern Recognition Letters, 2011,32 (10):1456-1467. (SCI: 000292236500005, EI: 20112114009105. IF:1.586).
[16] Jianhua Jia, Licheng Jiao, Xia Chang. Soft Spectral Clustering Ensemble Applied to Image Segmentation. Frontiers of Computer Science in China, Springer Press, 2011,5(1):66-78. (SCI:000292506200007, EI:20110913712647).
[17] Jianhua Jia, Xuan Xiao, Binxiang Liu. Similarity-based spectral clustering ensemble selection. Proceedings of the 9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2012).2012:1071-1074.
[18] Wu L, Gerard N P, Wyatt R, et al.CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5. Nature,1996,384:184-187[DOI].
[19] Cotch F, McAdam S N, Allsopp C E et al. Cytotoxic T cells in HIV2 seropositive Gambinas, Identification of a virus-specific MHC-restricted peptide epitope[J]. J Immunol, 1993; 151: 3361.
[20] http://so.med.wanfangdata.com.cn/ViewHTML/ConferencePaper_7127525.aspx.