

**Profil Poliprotein Virus Hepatitis C dan Analisa Kemungkinan
Molecular Mimicry dengan Protein *Homo sapiens* secara *In Silico***

SKRIPSI

oleh
Fachresa Riski Perdana
0710910022-91



JURUSAN BIOLOGI
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS BRAWIJAYA
MALANG
2011

**Profil Poliprotein Virus Hepatitis C dan Analisa Kemungkinan
Molecular Mimicry dengan Protein *Homo sapiens* secara *In Silico***

SKRIPSI

Sebagai salah satu syarat untuk memperoleh gelar
Sarjana Sains dalam bidang Biologi

oleh

FACHRESA RISKI PERDANA

0710910022-91



**JURUSAN BIOLOGI
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS BRAWIJAYA
MALANG
2011**

LEMBAR PENGESAHAN SKRIPSI

**Profil Poliprotein Virus Hepatitis C dan Analisa Kemungkinan
Molecular Mimicry dengan Protein *Homo sapiens* secara *In Silico***

oleh

**FACHRESA RISKI PERDANA
0710910022-91**

**Setelah dipertahankan di depan Majelis Penguji
pada tanggal 01 Agustus 2011
dan dinyatakan memenuhi syarat untuk memperoleh gelar
Sarjana Sains dalam bidang Biologi**

Pembimbing I

Pembimbing II

**Muhaimin Rifa'i Ph.D.Med.Sc
NIP. 19670525-199103-2-001**

**Widodo, SSi.,MSi. PhD.Med.Sc
NIP. 19730811-200003-1-002**

**Mengetahui,
Ketua Jurusan Biologi
Fakultas MIPA Universitas Brawijaya**

**Widodo, SSi., M.Si. PhD.Med.Sc
NIP. 19730811-200003-1-002**

LEMBAR PERNYATAAN

Saya yang bertanda tangan di bawah ini :

Nama : Fachresa Riski Perdana
NIM : 0710910022-91
Jurusan : Biologi
Penulis Skripsi Berjudul : Profil Poliprotein Virus
Hepatitis C dan Analisa Kemungkinan *Molecular Mimicry*
dengan Protein *Homo sapiens* secara *In Silico*

Dengan ini menyatakan bahwa :

1. Skripsi ini adalah benar-benar karya saya sendiri, dan bukan hasil plagiat dari karya orang lain. Karya-karya yang tercantum dalam Daftar Pustaka ini, semata-mata digunakan sebagai acuan/referensi.
2. Apabila kemudian hari diketahui bahwa isi skripsi merupakan hasil plagiat, maka saya bersedia menanggung segala resiko yang akan saya terima.

Demikian pernyataan ini saya buat dengan segala kesadaran.

Malang, 01 Agustus 2011
Yang menyatakan,

(Fachresa Riski Perdana)
NIM. 0710910022-91

PEDOMAN PENGGUNAAN SKRIPSI

Skripsi ini tidak dipublikasikan namun terbuka untuk umum dengan ketentuan hak cipta ada pada penulis. Daftar pustaka diperkenankan untuk dicatat, tetapi pengutipan hanya dapat dilakukan dengan seijin penulis dan harus disertai kebiasaan ilmiah untuk menyebutkannya.

UNIVERSITAS BRAWIJAYA



KATA PENGANTAR

Alhamdulillah Rabbil ‘Aalamiin, dengan ungkapan rasa syukur pada Allah Yang Maha Kuasa penulis dapat menyelesaikan penyusunan skripsi yang merupakan syarat untuk memperoleh gelar Sarjana Sains dalam bidang Biologi di Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Brawijaya Malang.

Pada kesempatan ini, penulis ingin menyampaikan ucapan terima kasih kepada :

1. **Kedua orang tua** atas segala doa, jasa, dan pengorbanan yang tak mungkin terbalas.
2. **Bapak Muhaimin Rifa’i, S.Si., M.Si. PhD. Med.Sc dan Bapak Widodo, S.Si., M.Si. PhD. Med.Sc** selaku Dosen Pembimbing yang telah mendampingi dan memberi pengarahan serta tambahan ilmu dan saran-saran yang berguna bagi penulis.
3. **Ibu Dr. Sri Widyarti, M.Si, Bapak Brian Rahardi S.Si., M.Sc , dan Bapak Dr. Suharjono, M.Si.** selaku Dosen Penguji yang telah memberi saran yang bermanfaat demi perbaikan penyusunan skripsi.
4. **Dra. Nunung Harijati, MS, Ph.D** selaku dosen pembimbing akademik yang telah mendampingi dan memberikan nasehat akademik
5. Keluarga besar tercinta yang telah memberikan doa, dukungan dan motivasi tiada henti.
6. Rekan seperjuangan “*Quorum Sensing Bio’ 2007*” dan segenap keluarga besar Jurusan Biologi FMIPA-UB atas bantuan dan kerja sama.
7. Teman-teman angkatan 2005, 2006, 2008, 2009 dan 2010 atas semangat dan kekeluargaan yang erat.

Penulisan skripsi ini merupakan upaya optimal penulis sebagai sarana terbaik dalam pengembangan ilmu pengetahuan. Saran dan kritik yang membangun sangat diharapkan untuk menjadikan karya ini semakin bermanfaat.

Malang, Agustus 2011

Penulis

Profil Poliprotein Virus Hepatitis C dan Analisa Kemungkinan *Molecular Mimicry* dengan Protein *Homo Sapiens* secara *In Silico*

Fachresa R.P, Muhaimin Rifa'I, Widodo
Jurusan Biologi, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Brawijaya, Malang.

ABSTRAK

Autoimun Hepatitis C masih menjadi permasalahan yang belum banyak ditemukan solusinya. Penyebab autoimun hepatitis C diduga karena *molecular mimicry*. Maka pada penelitian ini dilakukan *profiling* polyprotein pada Virus Hepatitis C dan mendeteksi *molecular mimicry* secara *in silico*. Tujuan dari penelitian ini adalah mengetahui profil poliprotein Virus Hepatitis C serta memprediksi dan mengetahui potensi munculnya *molecular mimicry* pada genom Virus Hepatitis C terhadap protein *Homo sapiens*. Metode yang digunakan dalam penelitian ini adalah dengan mendapatkan sekuen Virus Hepatitis C dari *International Comitte Taxonomy of Virus*, analisis *alignment* dan pohon kekerabatan dengan BioEdit dan MEGA 4, dan analisis *molecular mimicry* dengan BLAST. Polyprotein virus hepatitis C memiliki 17 *highly conserved region* dengan parameter 10 asam amino atau 30 nukleotida. Sepuluh daerah dari 17 sekuen asam amino terletak pada *functional region* dan 7 lainnya terletak pada *non functional region*. Dugaan mutasi pada polyprotein virus hepatitis C terdapat hampir pada seluruh region. Dugaan mutasi banyak terjadi pada *NS4 region*, *NS2 region*, dan *envelope region*. Infeksi virus hepatitis C diduga dapat menginduksi terjadinya *molecular mimicry* dimana dapat memunculkan autoantibodi pada *specific organ* (hepar) dan *non specific organ*.

Kata kunci: *In silico*, *Molecular mimicry*, Poliprotein, Virus Hepatitis C

Profile of Hepatitis C virus Polyprotein and Analysis of Possible Molecular Mimicry with *Homo Sapiens* Protein by In silico Methods

Fachresa R.P, Muhaimin Rifa'I, Widodo
Biology Departement, Mathematics and Science Faculty,
Brawijaya University, Malang.

ABSTRACT

Autoimmune Hepatitis C is still a problem. Autoimmune Hepatitis C may cause by of molecular mimicry as a main cause. The purpose of this study are to determine the profile of Hepatitis C Virus polyprotein as well as predicting and understanding the potential emergence of molecular mimicry in the genome of the Hepatitis C virus with *Homo sapiens* protein. To get the Hepatitis C virus sequences, it takes from the International Comitte Taxonomy of Viruses, alignment and phylogenic tree analysis with BioEdit and MEGA 4, and analysis of molecular mimicry with BLAST. Polyprotein hepatitis C virus has 17 highly conserved region with the parameters of 10 amino acids or 30 nucleotides. Ten regions of 17 amino acid sequences located on functional region and the other 7 are located on the non-functional regions. Allegations of mutations in hepatitis C virus polyprotein contained almost the entire region. Allegations of many mutations occur in NS4 region, NS2 region, and the envelope region. Suspected hepatitis C virus infection can induce the occurrence of molecular mimicry of autoantibodies which can result in specific organs (liver) and non-specific organs.

Keywords: In silico, Molecular mimicry, polyprotein, Hepatitis C Virus

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
HALAMAN PERNYATAAN	iii
ABSTRAK	iv
ABSTRACT	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR GAMBAR	viii
DAFTAR TABEL	ix
DAFTAR LAMPIRAN	x
DAFTAR ISTILAH	xi
PEDOMAN PENGGUNAAN SKRIPSI	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Permasalahan	2
1.3 Tujuan.....	2
1.4 Manfaat.....	2
BAB II TINJAUAN PUSTAKA	3
2.1 Tinjauan Umum Tentang Hepatitis C	3
2.1.1 Tinjauan Tentang Genome dan Antigenisitas Virus Hepatitis C.....	3
2.1.2 Tinjauan Tentang Respon Immunologis Terhadap Virus Hepatitis C.....	5
2.2 Tinjauan Umum Tentang Autoimun.....	7
2.3 Tinjauan Tentang Analisis <i>Alignment</i>	9
2.4 <i>Software</i> Bioinformatik	10
2.4.1 BioEdit	10
2.4.2 MEGA (<i>Phylogenetic and Molecular Evolutionary Analysis</i>).....	10
2.4.3 ICTVdb (<i>International Comitte on Taxonomy of Viruses Data Bases</i>)	11
2.5 Kerangka Konsep	11

BAB III METODE PENELITIAN	12
3.1 Waktu dan Tempat.....	12
3.2 Alat dan Bahan	13
3.2.1 Koleksi Data	12
3.2.2 Analisa Profil Poliprotein Virus Hepatitis C secara <i>In silico</i>	13
3.2.3 Analisis Potensi <i>Mimicry</i> Poliprotein Virus Hepatitis C terhadap Protein <i>Homo sapiens</i>	15
3.2.4 Kerangka Kerja	16
BAB IV HASIL DAN PEMBAHASAN	17
4.1 Genom Virus Hepatitis C	17
4.2 Pensejajaran Sekuen (<i>Alignment</i>), Analisis Daerah Terkonservasi (<i>Conserved Region</i>), dan Pohon Filogeni Genom Virus Hepatitis C.....	18
4.3 Analisis <i>Molecular Mimicry</i> dengan BLAST dan Bepired	24
BAB V PENUTUP	28
5.1 Kesimpulan	28
5.2 Saran	28
DAFTAR PUSTAKA	29
LAMPIRAN	32

DAFTAR GAMBAR

2.1	Struktur genome dari virus hepatitis C.....	4
2.2	Kerangka konsep penelitian	11
2.3	Kerangka kerja penelitian	14
3.1	Proses analisis <i>conserved region</i> dengan BioEdit	15
4.1	Pohon filogeni dari varian virus hepatitis C dengan MEGA 4	19
4.2	Perbandingan daerah yang terkonservasi dengan daerah yang diduga mengalami mutasi.	20
4.3	Hasil analisis epitop sel B dengan Bepipred dan sekuen masing- masing protein.....	23



DAFTAR TABEL

4.1 Genom virus Hepatitis C dan virus ParaInfluenza tipe 1 (sebagai <i>outgroup</i>)	16
4.2 Letak <i>conserved region</i> pada poliprotein	17
4.3 Protein <i>mimic</i> (Polyprotein HCV- <i>Homo sapiens</i>).....	21

UNIVERSITAS BRAWIJAYA



DAFTAR ISTILAH

Istilah

Keterangan

Alignment

Pensejajaran Nukleotida atau asam amino untuk menentukan homologi berdasarkan posisi sekuen

Capsid

Protein pembungkus atau pelindung virus

Envelope

Bagian terluar dari virus yang melindungi bagian capsid

Hypergammaglobulinaemia

Keadaan meningkatnya Imunoglobulin pada darah

In silico

Metode Analisis biologi yang dilakukan dengan perangkat komputer secara menyeluruh

Molecular mimicry

Suatu kelainan imunologis disebabkan adanya kesamaan struktur antara self protein dengan protein antigen dapat memicu respon imun

NS1

Menginisiasi Internal Ribosomes Entry Sites

NS2

Protein viral yang berfungsi untuk berikatan pada membran retikulum endoplasma

NS3 (peptidase S29) dan NS4

kompleks NS3 dan NS4 untuk pemutusan atau pembelahan antar protein (non struktural protein)

NS5A dan NS5B

Anti Interferon dari virus

Polyprotein

Protein dalam ukuran yang besar dan kemudian dipotong menjadi protein yang lebih kecil dengan fungsi biologis yang berbeda

RNA *dependen* RNA
Polymerase

Enzim yang mengkatalisis replikasi RNA dari template RNA

DAFTAR LAMPIRAN

1 Hasil analisis <i>conserved region</i> dengan <i>BioEdit</i>	32
2 <i>Alignment region</i> poliprotein Virus Hepatitis C	40

UNIVERSITAS BRAWIJAYA



BAB I

PENDAHULUAN

1.1 Latar belakang

Virus Hepatitis C dan Hepatitis B berperan dalam terjadinya 64% penyakit hati kronik. Virus Hepatitis C dan B merupakan penyebab terbanyak dikawasan Asia termasuk Indonesia. Infeksi ini dapat berlanjut menjadi sirosis hati. Infeksi virus Hepatitis C merupakan penyebab utama dari Hepatitis C dan masih menjadi penyebab utama dari hepatitis kronik di seluruh dunia (Hermawan, 2004). Autoimun adalah respon sistem imun yang terjadi terhadap antigen jaringan tubuh sendiri yang disebabkan kegagalan mekanisme secara normal yang berperan untuk mempertahankan *self tolerance* terhadap sel B, sel T, atau keduanya. Autoimunitas dapat terjadi karena *self*-antigen yang dapat menimbulkan aktivasi, proliferasi, dan diferensiasi sel T autoreaktif menjadi sel efektor yang menimbulkan kerusakan jaringan dan organ (Baratawidjaja, 2006). Salah satu autoimun yang berhubungan metabolisme tubuh yaitu autoimun hepatitis.

Autoimun hepatitis merupakan suatu kelainan imunologis yang menyebabkan inflamasi *hepar* dan peningkatan IgG (*Hypergammaglobulinaemia*). Penyakit autoimun hepatitis hingga saat ini belum diketahui penyebabnya secara pasti. Penyebab yang diduga kuat sebagai penyebab autoimun hepatitis adalah infeksi virus hepatitis dan obat- obatan tertentu yang digunakan.

Analisis secara bioinformatik merupakan salah satu pendekatan yang dapat dilakukan, di mana peneliti dapat memprediksi penyebab dan *treatment* suatu kelainan atau *disorder*. Autoimun hepatitis juga dapat menurunkan aktivitas dari sel T regulator di mana peran sel T regulator sangat penting untuk meregulasi hipereaktivitas sel- sel imunokompeten. Dengan dilakukannya analisis secara bioinformatik, dapat diketahui profil genomik Virus Hepatitis C sebagai sarana informasi desain vaksin serta dapat diprediksi penyebab autoimun hepatitis. Oleh karena itu, analisis bioinformatik (*in silico*) tentang profil poliprotein Virus Hepatitis C dan penyebab autoimun hepatitis perlu dilakukan untuk

mengambil tindakan pengobatan yang tepat serta pencegahan yang sesuai.

1.2 Rumusan Permasalahan

Permasalahan yang terdapat dalam penelitian ini adalah:

1. Bagaimana profil Poliprotein Virus Hepatitis C?
2. Bagaimana kemungkinan munculnya protein *mimic* pada *Homo sapiens*?

1.3 Tujuan

Tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Mengetahui profil Poliprotein Virus Hepatitis C.
2. Memprediksi dan mengetahui kemungkinan munculnya protein *mimic* pada *Homo sapiens*.

1.4 Manfaat

Manfaat penelitian ini adalah dapat diprediksinya penyebab autoimun hepatitis, serta profil Genom dan poliprotein Virus hepatitis C, sehingga dari penyebab tersebut dapat dirancang metoda pengobatan baru (obat ataupun vaksin) yang spesifik terhadap masing- masing varian virus sebagai bentuk pencegahan dan pengobatan autoimun hepatitis.

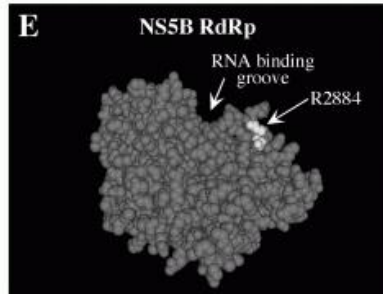
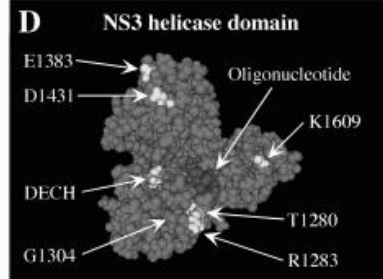
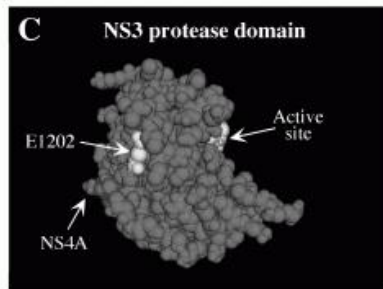
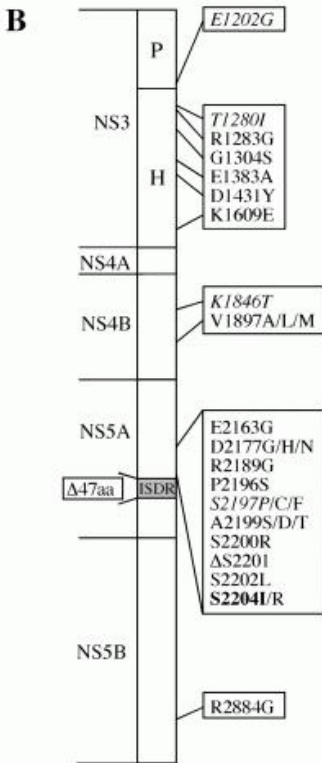
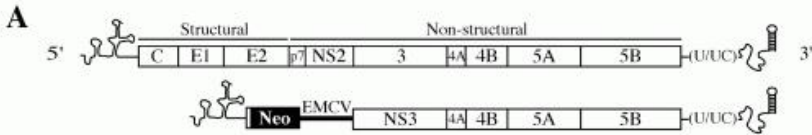
BAB II TINJAUAN PUSTAKA

2.1 Hepatitis C

Hepatitis C adalah penyakit yang disebabkan oleh virus hepatitis C. Infeksi Virus Hepatitis C dapat menyebabkan peradangan hepar (hepatitis) yang biasanya asimtomatik, tetapi hepatitis kronik dapat menyebabkan sirosis dan kanker hepar. Virus Hepatitis C dapat menyebar melalui kontak darah (transfusi darah) dan kontak seksual. Hepatitis B dan C merupakan masalah kesehatan global yang sering dihadapi. Kedua virus tersebut dapat berakibat pada hepatitis kronik, sirosis dan *hepatoselular carcinoma* (Miyasaki, 2009).

2.1.1 Genome dan Antigenisitas Virus Hepatitis C

Virus Hepatitis C adalah virus bertipe RNA yang memiliki *envelope* berdiameter 50 nm, terklasifikasi pada Genus Hepacivirus dan Famili Flaviridae. Virionnya menginfeksi manusia. Genome dari Virus Hepatitis C sangat mudah termutasi karena Virus Hepatitis C merupakan virus dengan tipe RNA virus. Karena Virus Hepatitis C juga memiliki sifat cepat termutasi, maka Virus Hepatitis C mudah untuk *escaping* dari sistem imun yang ada dalam tubuh. Mutasi yang terjadi pada genome Virus Hepatitis C tidak terjadi pada seluruh bagian genome, melainkan terjadi pada hipervariabel region yang mengkode protein *envelope* (berlokasi dekat N-terminus dari E2). Peneliti memprediksi bahwa pada bagian ini yang memiliki antigenisitas tinggi. Sejauh ini genotip yang dimiliki diketahui terdapat 6 genotip (WHO, 2010).



Gambar 2.1 Struktur genome dari Hepatitis C Virus

A dan B menunjukkan penempatan *site* pada genom virus Hepatitis Letak poliprotein C) NS3 protease D) NS3 helicase E) NS5B RNA dependen RNA polimerase (Blight dan Norgard, 2006).

Virus Hepatitis C termasuk dalam famili *Flaviridae* memiliki struktur dan karakteristik yang spesifik. Kelompok ini memiliki *envelope* pada lipid bilayer di mana dua atau lebih *envelop* protein yang tertancap. *Envelop* mengelilingi nukleokapsid, di mana terdiri dari beberapa salinan (*copy*) dari protein kecil (dari *Core protein* dan mengandung RNA genom). Genom *flaviridae* merupakan untai positif (*positive strand*) molekul RNA yang berukuran 9,6 hingga 12,3 ribu nukleotida, dengan daerah *open reading frame* (ORF) yang mengkode poliprotein sebanyak 3000 asam amino bahkan lebih (Chevaliez dan Pawlotsky, 2006).

Flaviridae berikatan dengan satu atau lebih reseptor sel, yaitu reseptor kompleks yang menginisiasi endositosis virus. Masuknya viral ini dikontrol oleh glikoprotein pada membran virion yang menginisiasi perubahan konformasi yang memediasi fusi pada fusi protein *class II*. Fusi dari *virion envelope* dengan membran sel yang meneruskan nukleokapsid ke sitoplasma. Setelah peleburan (dekapsidasi) *envelop*, translasi viral genom akan terjadi di sitoplasma, yang mengarahkan pada sintesis poliprotein prekursor yang akan terbagi menjadi protein struktural dan non struktural dari genom virus. Viral tersebut bereplikasi, dan kemudian berasosisasi dengan membran sel *host*, setelah virion tersebut *terpackage* dengan vesikel pada *Golgi apparatus*, maka virion tersebut telah matang dan akan dikeluarkan ke ekstrasel melalui eksositosis (Chevaliez dan Pawlotsky, 2006).

2.1.2 Respon Immunologis terhadap Virus Hepatitis C

Virus mempunyai sifat-sifat khusus di antaranya: dapat menginfeksi jaringan tanpa menimbulkan respon inflamasi, dapat berkembang biak dalam sel *host* tanpa merusaknya, kadang mengganggu fungsi khusus sel yang terinfeksi tanpa merusaknya secara nyata, dan kadang virus merusak sel dan mengganggu perkembangan sel, kemudian menghilang dari tubuh. Respon imun terhadap infeksi virus melibatkan respon non-spesifik maupun spesifik. Terdapat 2 mekanisme utama respon non spesifik terhadap virus, yakni: infeksi virus merangsang IFN (*Interferon*) oleh sel-sel terinfeksi, di mana IFN berfungsi menghambat replikasi virus dan sel NK (*Natural Killer*) melisiskan berbagai jenis sel terinfeksi virus. Sel

NK mampu melisis sel terinfeksi virus walaupun virus menghambat presentasi antigen dan ekspresi MHC I (*Major Histocompatibility Complex*), karena sel NK cenderung diaktivasi oleh sel sasaran yang MHC- negatif (Kresno, 2001).

Untuk membatasi penyebaran virus dan mencegah reinfeksi, sistem imun tubuh harus mampu menghambat masuknya virion ke dalam sel dan memusnahkan sel yang terinfeksi. Antibodi spesifik mempunyai peran penting pada awal terjadinya infeksi, di mana antibodi ini dapat menetralkan antigen virus dan melawan virus sitopatik yang dilepaskan oleh sel yang mengalami lisis. Peran antibodi dalam menetralkan virus terutama efektif untuk virus yang bebas atau virus dalam sirkulasi. Proses netralisasi virus dapat dilakukan dengan beberapa cara, di antaranya dengan cara menghambat perlekatan virus pada reseptor yang terdapat pada permukaan sel, sehingga virus tidak dapat menembus membrane sel, dengan demikian replikasi virus dapat dicegah. Antibodi dapat juga menghancurkan virus dengan cara aktivasi komplemen melalui jalur klasik atau menyebabkan agregasi virus sehingga mudah difagositosis dan dihancurkan. Antibodi dapat mencegah penyebaran virus yang dikeluarkan dari sel yang telah hancur. Tetapi seringkali antibodi tidak cukup mampu untuk mengendalikan virus yang telah mengubah struktur antigennya dan yang melepaskan diri (*budding off*) melalui membran sel sebagai partikel yang infeksius, sehingga virus dapat menyebar ke dalam sel yang berdekatan secara langsung. Walaupun tidak cukup mampu menetralkan virus secara langsung, antibodi dapat berfungsi dalam reaksi ADCC (*Antibody Dependent Cellular Cytotoxicity*) (Kresno, 2001).

Infeksi virus Hepatitis C merangsang pembentukan antibodi terhadap protein virus yang nonstruktural yang disebut anti-HCV. Anti-HCV muncul lama setelah infeksi, bahkan kadang baru muncul 16-24 minggu setelah peningkatan kadar enzim. Pada infeksi akut yang umumnya sembuh sendiri, anti-HCV juga menghilang dari sirkulasi, sedangkan pada infeksi kronik anti-HCV menetap pada sirkulasi dalam jangka waktu lama dengan kadar tinggi (Kresno, 2001).

2.2 Autoimun

Autoimun merupakan salah satu penyakit disregulasi sistem imunitas dan merupakan penyebab penting dalam penyakit yang diderita manusia di beberapa negara. Kata autoimun sering diartikan bahwa kerusakan dan ketidaknormalan dari beberapa penyakit yang melibatkan sistem imunitas dengan kerusakan jaringan karena sistem imun itu sendiri (Abbas dkk, 2009). Beberapa konsep yang dikembangkan hingga saat ini dalam autoimunitas (Baratawidjaja, 2006):

1. Teori *sequestered antigen* atau *hidden antigen*

Sequestered atau *hidden antigen* adalah antigen yang karena secara anatomi tidak pernah berhubungan dengan sistem imun misalnya antigen sperma, lensa mata, dan saraf pusat. Bila bagian tersebut rusak atau terluka, dapat timbul penyakit autoimun.

2. *Molecular mimicry*

Banyak infeksi yang menunjukkan hubungan dengan penyakit autoimun tertentu. Beberapa antigen memiliki epitop yang sama dengan antigen sel sendiri. Respon imun yang muncul terhadap antigen tersebut dapat merangsang sel T dan kemudian juga merangsang sel B untuk membentuk autoantibodi. Infeksi bakteri dan virus dapat berkontribusi dalam terjadinya gangguan autoimunitas (Baratawidjaja, 2006).

Molecular mimicry merupakan salah satu penyebab autoimunitas. Hal ini disebabkan adanya kesamaan struktur antara *self* protein dengan protein antigen dapat memicu respon imunologis. *Self Peptide* dengan konsentrasi rendah dapat mengakibatkan reaksi silang dengan peptida antigen yang memiliki struktur serupa (*mimic*). Hal ini mengakibatkan perpindahan atau ekspansi populasi sel T yang responsif dan dapat mengenali *self peptide*. Pada keadaan ini, toleransi sel T akan terganggu terhadap peptida suatu sel, maka inflamasi akan berlanjut pada presentasi peptida lainnya dan respon imun akan meluas sehingga menghasilkan kerusakan jaringan lokal. Kejadian ini disebut *epitope spreading*.

3. Kegagalan autoregulasi

Regulasi sistem imun berfungsi untuk mempertahankan homeostasis. Gangguan terjadi pada presentasi antigen, infeksi yang dapat meningkatkan respon MHC, kadar sitokin yang rendah, dan gangguan terhadap respon IL-2. Kegagalan pada sel T *regulator* atau sel T *supresor* dapat menimbulkan autimunitas (Baratawidjaja, 2006).

4. Obat- obatan

Antigen asing dapat diikat oleh permukaan sel dan menimbulkan reaksi kimia dengan antigen permukaan sel tersebut yang dapat mengubah imunogenisitasnya. Mekanisme terjadinya reaksi autoimun karena obat belum jelas hingga saat ini. Banyak reaksi yang ditampilkan ketika mengkonsumsi suatu obat seperti munculnya *Antinuclear Antibodi* (ANA) dan antibodi tersebut akan menghilang ketika pemakaian obat dihentikan, tetapi tidak sedikit obat masih banyak tersimpan dalam tubuh (Baratawidjaja, 2006).

5. Faktor keturunan

Penyakit autoimun mempunyai persamaan predisposisi genetik. Meskipun sudah diketahui adanya kecenderungan terjadi penyakit pada keluarga, tetapi mekanisme penurunannya pada keturunan selanjutnya belum diketahui. Hal ini dapat dibuktikan dengan adanya keterkaitan antara penyakit dengan HLA (*Human Leukocyte Antigen*). *Haplotipe* HLA merupakan resiko relatif untuk penyakit autoimun tertentu (Baratawidjaja, 2006).

Pada penyakit Autoimun hepatitis terjadi kerusakan jaringan hepar yang kronis. Pada kasus autoimun hepatitis terdapat antibodi yang menyerang organ dan jaringan hepar. Bila keadaan memburuk akan menyebabkan peradangan yang akan berkembang menjadi sirosis hepar dan terjadi kegagalan fungsi hepar. Autoimun hepatitis diduga dapat terjadi karena induksi obat- obatan dan bahan kimia serta infeksi virus. Hingga saat ini belum ada sumber pasti yang menyebutkan tentang etiologi atau penyebab autoimun hepatitis (Univesity of Virginia. 2007).

Hepatitis merupakan inflamasi yang terjadi pada hepar, yang mengakibatkan kerusakan sel hepar. Pada autoimun hepatitis, sistem imun pada tubuh penderita akan merusak sel hepar. Inflamasi kronik hepatitis ini masih belum diketahui penyebabnya secara pasti. Kondisi ini berasosiasi dengan kelainan yang biasa disebut *hypergammaglobulinaemia*. *Hypergammaglobulinaemia* adalah suatu kelainan di mana terdapat protein antibodi yang terdapat di dalam darah. Infeksi yang terjadi pada tubuh juga dapat menyebabkan *hypergammaglobulinaemia* (University of Virginia, 2007).

2.3 Analisis Alignment

Dalam bioinformatik, *sequence alignment* merupakan salah satu cara mengatur (*arranging*) sekuen DNA, RNA, atau protein untuk identifikasi kesamaan wilayah (*region*) yang berpengaruh pada struktural fungsi dan evolusi di antara sekuen tersebut. *Aligned sequence* dari residu nukleotida atau asam amino secara khusus direpresentasikan sebagai baris dalam sebuah matriks. *Gaps* diinsersikan di antara residu tersebut sehingga karakter yang identik atau sama diatur (disejajarkan) dalam *successive column*. *Comparative analysis* dari sekuen DNA dan asam amino merupakan komponen yang penting dalam penelitian di bidang Biologi. *Sequence alignment* telah membantu dalam studi evolusi molekular, *RNA folding*, regulasi gen, dan hubungan *protein structure-function*. Meskipun, perbandingan dari *pairwise sequence* telah terbukti bermanfaat, beberapa kesamaan penting secara biologi mungkin dideteksi dengan *aligning a set of sequence*. Demikian juga, pola atau motif dari set fungsional protein tersebut mungkin secara nyata dari analisis *multiple alignment* sekuen tersebut (Lipman, dkk., 1989).

Multiple sequence alignment merupakan prosedur untuk merubah *sequence of unequal length* ke *sequence of equal length* dengan menduga *placement of gaps* dengan tujuan untuk menduga homologi di antara karakter-karakter tersebut. Insersi dan delesi (indels) dilakukan dalam variasi cara selama *multiple sequence alignment* dan rekonstruksi filogenetik (Ogden dan Rosenberg, 2006).

Dua sekuen dalam sebuah *alignment share a common ancestor*, *mismatches* dapat diinterpretasikan sebagai mutasi titik (*point mutation*) dan *gaps* sebagai indels (insersi atau delesi). Pada *sequence alignment* protein, derajat kesamaan di antara asam amino berada pada posisi tertentu dalam sekuen sehingga dapat diinterpretasikan sebagai *rough measure* dari *conserved particular region* atau motif sekuen (*sequence motif*) di antara *lineage*. Ketidakadaan substitusi atau keberadaan dari *conservative substitution* dalam region tertentu dari sekuen menegaskan bahwa region ini memiliki struktur dan fungsi yang penting. Meskipun basa nukleotida DNA dan RNA lebih sama satu dengan yang lain daripada asam amino, konservasi dari pasangan basa dapat mengindikasikan peran yang sama pada fungsi dan strukturnya. Sekuen yang pendek atau sangat mirip dapat di-*aligned* dengan menggunakan tangan. Meskipun demikian, masalah utamanya adalah membutuhkan *alignment* yang panjang, *highly variable* atau jumlah sekuen ekstrim dimana hal ini tidak dapat dilakukan dengan tangan manusia. Sementara itu, pengetahuan manusia diaplikasikan dalam pembuatan algoritma untuk dapat menghasilkan *highly-quality sequence alignment*.

2.4 Software Bioinformatik

2.4.1. BioEdit

BioEdit adalah salah satu alat atau software untuk mengedit dan pensejajaran (*alignment*) sekuens. Program ini sering digunakan dalam membuat atau merancang alignment dari data- data yang didapatkan dari database serta dapat di atur ulang sesuai dengan keinginan pengguna. BioEdit juga menyediakan link untuk ditujukan kepada web server software lainnya untuk membantu pengambilan data maupun analisis (Meine, 2009).

2.4.2 MEGA (*Phylogenetic and Molecular Evolutionary Analysis*)

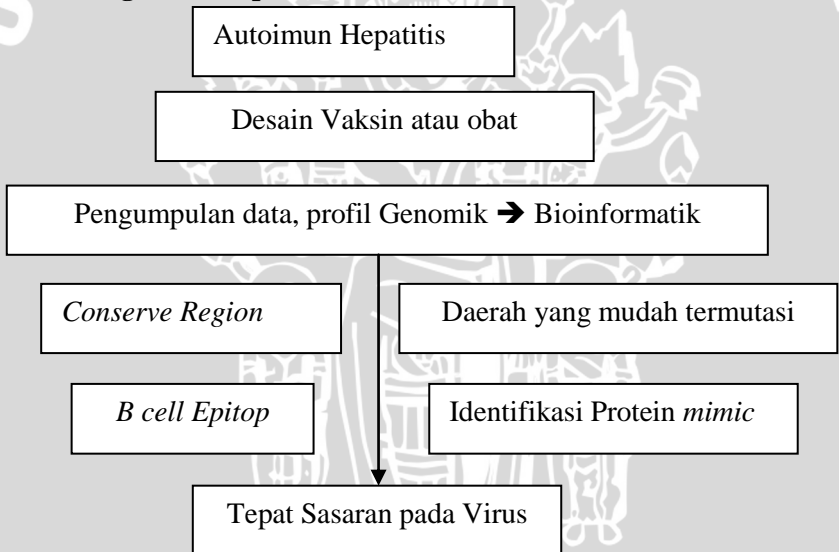
MEGA merupakan alat atau software yang memiliki integrasi untuk melakukan *sequence alignment* secara manual dan otomatis, mendesain pohon filogenetik dari data yang telah didapatkan, sumber

data base (*integrated*), memprediksi laju evolusi secara molekuler, dan membuktikan hipotesis evolusioner (MEGA, 2010).

2.4.3 ICTVdB (*International Comitte on Taxonomy of Viruses Data Bases*)

ICTVdB merupakan indeks dari virus atau *database* virus beserta deskripsi dan taksonomi. ICTVdB merupakan bentukan dari *The Seventh Report of the International Committee on Taxonomy of Viruses*, van Regenmortel *et al.* Academic Press (2000) dan data para peneliti tentang virus. Virus database ini dapat langsung memiliki *link* yang dihubungkan dengan NCBI sebagai *gene* dan protein databank (International Committee on Taxonomy of Viruses, 2002).

2.5 Kerangka Konsep



Gambar 2.2 Kerangka konsep penelitian: Desain vaksin untuk autoimun hepatitis dilakukan dengan mengumpulkan data genomik untuk memperoleh informasi sehingga dapat dibuat vaksin atau obat yang tepat sasaran.

BAB III

METODE PENELITIAN

3.1 Waktu dan Tempat

Penelitian “Profil Poliprotein Virus Hepatitis C dan Analisa Kemungkinan *Molecular Mimicry* dengan Protein *Homo sapiens* secara *In Silico*” dilaksanakan pada bulan Desember 2010- Mei 2011, bertempat di Laboratorium Biologi Informatika dan Komputasi, Jurusan Biologi, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Brawijaya, Malang.

3.2 Alat dan bahan

Alat dan bahan yang digunakan dalam penelitian ini adalah seperangkat komputer, jaringan atau koneksi internet, *Database*, *Software*, *data mining* dari *gene* atau *protein bank*.

3.2.1 Database

Sumber data (*database*) yang digunakan dalam protein ini adalah pada ncbi.nlm.nih.gov sebagai sumber sekuen dan ICTV *Viral Taxonomy Database* merupakan sumber informasi genotip virus Hepatitis C.

3.2.2 Software

Perangkat lunak (*software*) yang digunakan dalam penelitian ini yaitu MEGA *Phylogenetic and Molecular Evolutionary Analysis*, BioEdit *Sequence Alignment Editor*, BLASTp (NCBI) dan Immunepitope .

3.3 Cara kerja

3.3.1. Koleksi Data

Database genom Virus Hepatitis C didapatkan dari ICTVdB (*International Comitte Taxonomy of Virus database*) yang dihubungkan dengan *Link* ke situs *genebank* NCBI (*National Center of Biotechnology Information*). Sekuen nukleotida diunduh dan disimpan dalam format Fasta. Untuk mendapatkan sekuen asam

amino, dipilih *link* CDS pada halaman informasi genom tersebut. Sekuen asam amino diunduh dan disimpan dalam format Fasta.

3.3.2. Analisis Profil Poliprotein Virus Hepatitis C secara *In Silico*

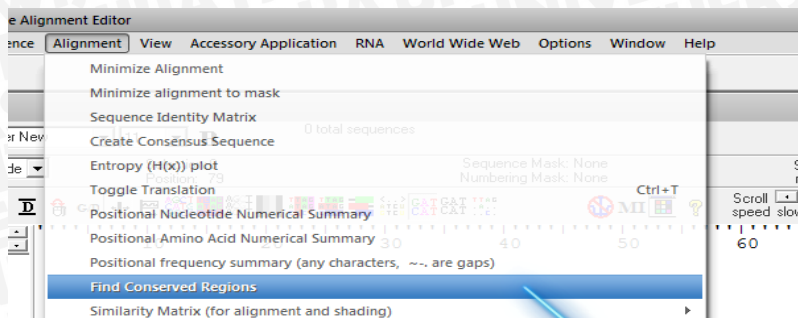
a. Pohon filogeni dan Homologi Antar Genom Virus Hepatitis C

Pembuatan pohon filogeni dan homologi antar genom virus Hepatitis C dilakukan dengan mensejajarkan sekuen (*alignment*) nukleotida dan asam amino masing- masing tipe dan ditambahkan 1 genome di luar takson Virus Hepatitis C sebagai *outgroup*. Analisis *alignment* ini menggunakan software MEGA 4 (*Molecular Evolutionary Genetics Analysis*). Setelah dilakukan *alignment*, dibuat pohon filogeni dari variabel genom tersebut untuk memvisualisasikan hasil analisis *alignment* tersebut. *Alignment* MEGA menggunakan ClustalW yang dikembangkan oleh Thompson dkk (1994) yang dapat mengalign Nukleotida dan Asam Amino. Analisis pohon filogeni dengan menggunakan *Neighbour Joining* yang menggunakan nilai perbandingan dua sekuen dan satu sekuen dengan seluruh sekuen yang dikembangkan oleh Yang dan Kumar (1996). Nilai yang tertera pada bagan merupakan skala jarak evolusioner yang dihitung menggunakan metode koreksi Poisson.

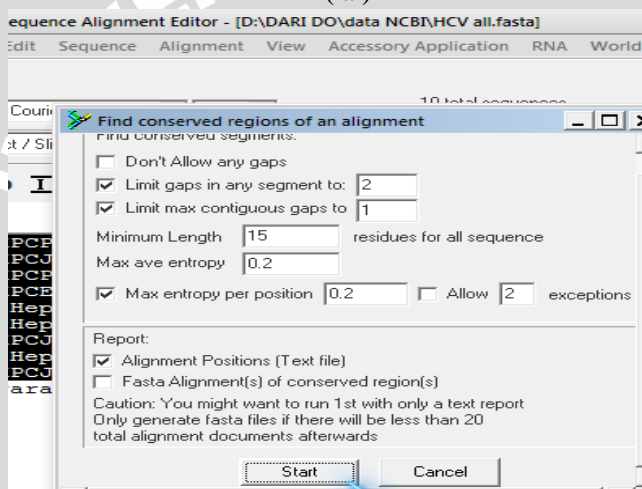
b. Analisis Daerah Terkonservasi (*Conserved Region*) dengan BioEdit

Analisis *Conserved Region* dilakukan setelah mensejajarkan (*alignment*) sekuen poliprotein. *Alignment* dilakukan dengan menggunakan program BioEdit dengan ClustalW yang algoritma digunakan oleh Thompson dkk (1994). Analisis daerah terkonservasi dapat dilakukan dengan cara

1. Dipilih menu *alignment*
2. Dipilih menu *find conserved region*
3. Setelah muncul kotak dialog,
4. Ditentukan parameter panajng sekuen
5. Dipilih *start* untuk memulai proses analisis



(a)



(b)

Gambar 3.1 Proses analisis *Conserved Region* dengan BioEdit

- a) Memilih menu *alignment* pada menu badan memilih *find conserve region*
- b) menentukan parameter panjang sekuen yang dianggap *conserve*

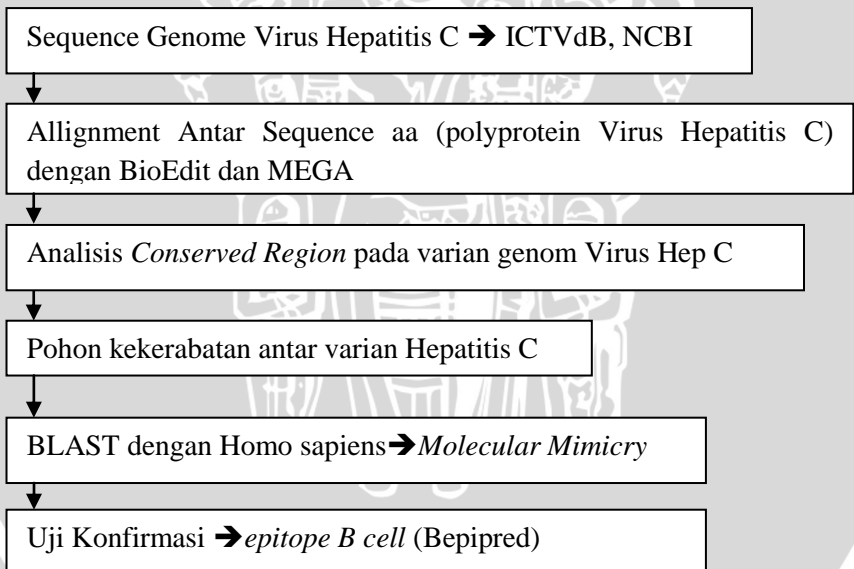
Setelah muncul hasil tersebut, kemudian data disesuaikan dengan region pada masing- masing varian virus Hepatitis C. Sehingga dapat ditentukan region yang terkonservasi pada seluruh varian Virus Hepatitis C.

3.3.3 Analisis Potensi *Mimicry* Poliprotein Virus Hepatitis C terhadap Protein *Homo sapiens*

Poliprotein Virus Hepatitis C didapatkan dari NCBI (telah tertera pada profil organisme) kemudian dipilih salah satu nama poliprotein yang dibentuk oleh Virus Hepatitis C untuk lebih spesifik sekuen yang muncul hanya untuk bagian salah satu protein yang dibentuk. Kemudian dipilih BLAST pada *menu bar*. Setelah masuk jendela konfigurasi BLAST, dipilih organisme yang akan dibandingkan yaitu *Homo sapiens*.

Protein yang telah didapatkan dan diduga *mimic* dengan protein *Homo sapiens* diuji konfirmasi dengan memprediksi epitope Sel B dengan menggunakan Bepipred dengan metode Larsen dan Nielsen (2006) untuk memastikan bahwa protein tersebut dapat memicu reaksi dari sel B untuk berproliferasi dan menghasilkan antibodi.

3.4 Kerangka kerja



Gambar 2.3 Bagan kerangka kerja penelitian

BAB IV HASIL DAN PEMBAHASAN

4.1 Genom Virus Hepatitis C

Data genom Virus Hepatitis C diambil dari ICTVdB (*International Comitte Taxonomy of Virus Data Base*). Data yang didapatkan, terdapat 9 Genom Virus Hepatitis C seperti tabel dibawah ini

Tabel 4.1. Genom Virus Hepatitis C dan Virus ParaInfluenza tipe 1 (sebagai *outgroup*)

No	Genom virus Hepatitis C	Accession Number (Nucleotide-Amino Acid)	Author
1	1A	M62321.1- AAA45676.1	Choo,Q.-L., Richman,K., Han,J.H., Berger,K., Lee,C., Dong,C., Gallegos,C., Coit,D., Medina-Selby,A., Barr,P.J., Weiner,A., Bradley,D.W., Kuo,G. and Houghton,M.
2	1B	D90208.1- BAA14233.1	Kato,N., Hijikata,M., Ootsuyama,Y., Nakagawa,M., Ohkoshi,S. and Shimotohno.K.
3	2A	D00944.1- BAA00792.1	Okamoto,H., Okada,S., Sugiyama,Y., Kurai,K., Iizuka,H., Machida,A., Miyakawa,Y. and Mayumi,M.
4	3A	D17763.1- BAA04609.1	Sakamoto,M., Akahane,Y., Tsuda,F., Tanaka,T., Woodfield,D.G. and Okamoto,H.
5	4A	Y11604.1- CAA72338.1	Chamberlain,R.W., Adams,N., Saeed,A.A., Simmonds,P. and Elliott,R.M.
6	5A	Y13184- CAA73640.1	Chamberlain,R.W., Adams,N.J., Taylor,L.A., Simmonds,P. and Elliott,R.M.
7	6A	Y12083.1- CAA72801.1	Adams,N.J., Chamberlain,R.W., Taylor,L.A., Davidson,F., Lin,C.K., Elliott,R.M. and Simmonds,P.
8	10	D63821.1- BAA09890.1	Tokita,H., Okamoto,H., Iizuka,H., Kishimoto,J., Tsuda,F., Lesmana,L.A., Miyakawa,Y. and Mayumi,M.
9	11	D63822.1- BAA09891.1	Tokita,H., Okamoto,H., Iizuka,H., Kishimoto,J., Tsuda,F., Lesmana,L.A., Miyakawa,Y. and Mayumi,M.
10	Human ParaInfluenza Virus 1	M22347.1- AAA46800.1	Merson,J.R., Hull,R.A., Estes,M.K. and Kasel,J.A.

Pada tabel tersebut terdapat data Virus ParaInfluenza tipe 1 sebagai pembanding untuk dilakukan analisis kekerabatan dan daerah yang terkonservasi antar varian Virus Hepatitis C. Berdasarkan (Chevaliez dan Pawlotsky, 2006) terdapat 7 varian genom dari Virus Hepatitis C, sedangkan yang didapatkan pada ICTVdB terdapat 9 varian.

4.2 Pensejajaran sekuen (*Alignment*), analisis daerah Terkonservasi (*Conserved Region*), dan Pohon Filogeni Genom Virus Hepatitis C

Pensejajaran sekuen nukleotida berfungsi untuk menentukan daerah yang terkonservasi dari genom RNA pada virus Hepatitis C. Hal ini dapat dijadikan informasi untuk merancang vaksin, dimana dapat ditentukan letak daerah yang memiliki kesamaan struktur dan komposisi basa nukleotida, sehingga memiliki keefektifan untuk digunakan bagi semua varian.

Berdasarkan hasil analisis *alignment* dan *Conserved region* (lampiran 1), didapatkan 17 *conserved region* sebagai berikut

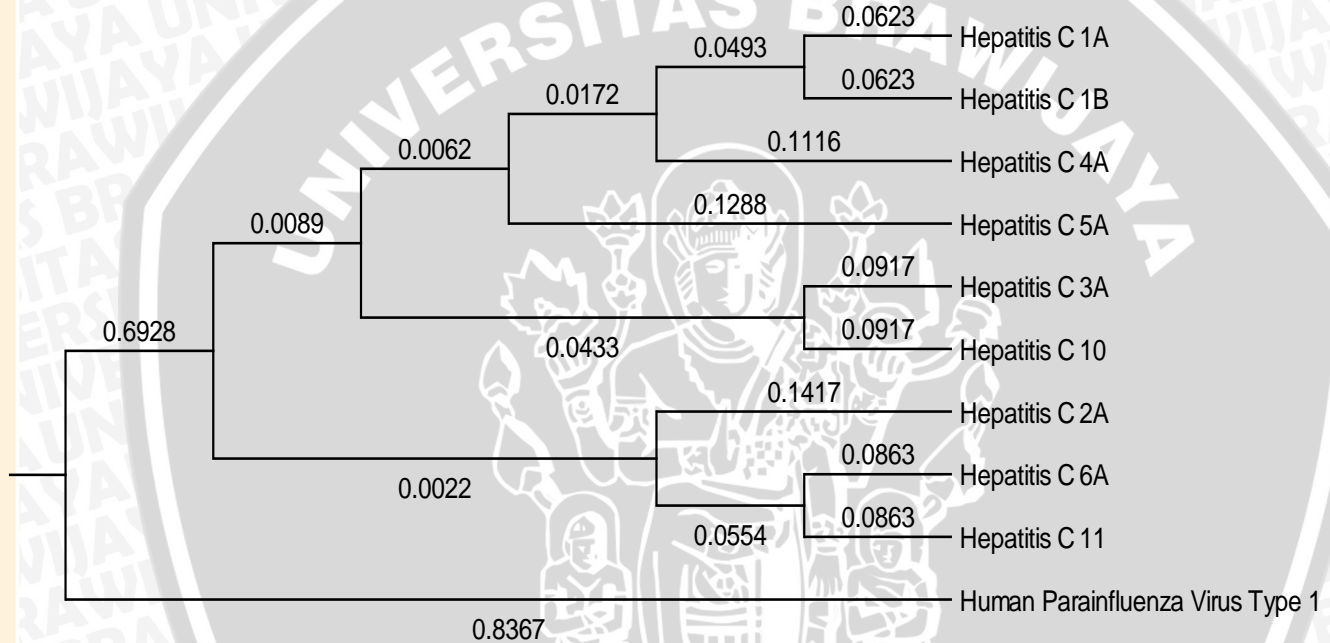
Tabel 4.2. Letak *Conserved Region* pada Poliprotein

POSISI	Jenis Poliprotein
Posisi 21 - 35	Capsid
Posisi 92 - 101	Capsid
Posisi 504 – 517	NS1
Posisi 588 – 597	NS1
Posisi 1258 – 1268	DEAD-like helicases superfamily
Posisi 1337 – 1347	DEAD-like helicases superfamily
Posisi 1393 – 1409	Helicase superfamily c-terminal domain (Helic C)
Posisi 1452 – 1462	Diantara NS3- NS4 (UTR)
Posisi 1627 – 1636	Diantara NS3- NS4 (UTR)
Posisi 1777 – 1793	NS4B
Posisi 1921 – 1935	Diantara NS4B- NS5A (UTR)
Posisi 2401 – 2415	Diantara NS5A 1B- RNA Dependen RNA Polimerase (UTR)

Posisi 2591 – 2600	RNA Dependen RNA Polimerase
Posisi 2672 – 2682	RNA Dependen RNA Polimerase
Posisi 2797 – 2807	RNA Dependen RNA Polimerase
Posisi 2819 – 2828	> RNA Dependen RNA Polimerase (UTR)
Posisi 2859 – 2868	> RNA Dependen RNA Polimerase (UTR)

Parameter diatur panjang sekuen untuk analisis *conserved region* yaitu 10 asam amino = 30 basa nukleotida. Hal ini bertujuan untuk mendapatkan region atau daerah yang *Highly Conserved*. Karena jika perancangan vaksin berdasarkan daerah yang *Highly Conserved*, akan dapat meningkatkan muatan antigenik pada epitope vaksin (Jones dkk, 2010).

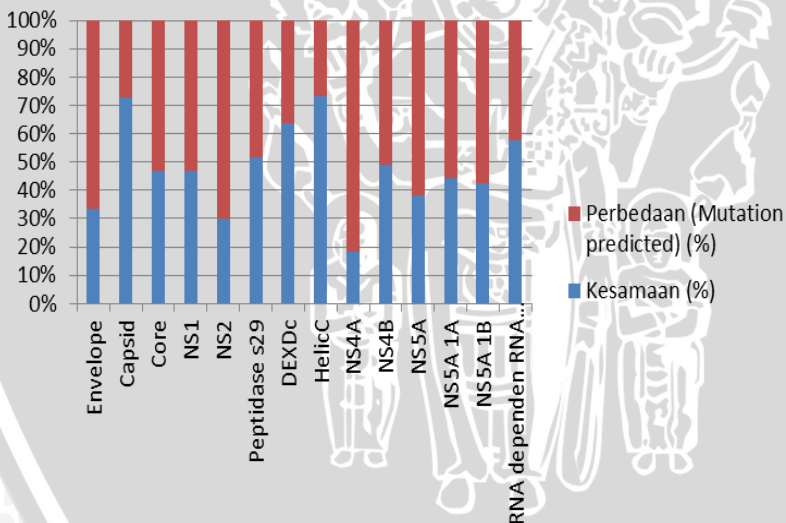
Berdasarkan data hasil analisis yang didapatkan, terdapat 11 *functional region* dan 6 *non functional region*. Hal ini menunjukkan terdapat *Highly Conserved Region* yang berada pada *functional region*. Keanekaragaman Genetis pada Virus Hepatitis C ini berkisar 60- 70 %, meskipun virus Hepatitis C ini mudah termutasi. Karena Virus Hepatitis C juga memiliki sifat cepat termutasi, maka Virus Hepatitis C mudah untuk *escaping* dari sistem imun yang ada dalam tubuh. Berdasarkan hasil yang didapatkan, variasi genetik pada genom virus hepatitis C, tidak hanya terdapat pada bagian *envelope*, tetapi juga terjadi pada bagian yang lain. Bagian yang tidak banyak terkonservasi adalah bagian *envelope*, NS2, dan NS3 yang termasuk bagian peptidase 29. Dari hal tersebut dapat dikatakan bahwa banyaknya variasi genetik yang terjadi pada daerah tersebut. maka *site* itulah terjadinya mutasi. Hasil yang didapatkan sesuai dengan pernyataan WHO (2010) dimana mutasi yang terjadi pada genom Virus Hepatitis C tidak terjadi pada seluruh bagian genom, melainkan terjadi pada hipervariabel region.



Gambar 4.1 Pohon filogeni dari varian virus Hepatitis C dengan MEGA 4

Dari pohon filogeni yang dihasilkan dari hasil analisis *alignment* dapat diartikan bahwa varian yang ditunjukkan urutan angka, bukan menunjukkan kekerabatan yang dekat, tetapi berdasarkan waktu penemuan Genom Virus Hepatitis C tersebut. Dari Pohon kekerabatan tersebut dapat diketahui keragaman ekspresi poliprotein Virus Hepatitis C. Hal ini berkaitan dengan hasil analisis *Conserved region* yang terdapat beberapa *site* yang tidak terkonservasi, atau dengan kata lain *site* tersebut mudah sekali termutasi.

Sebagai analisis tambahan dilakukan analisis *sequence alignment* pada masing- masing protein Virus Hepatitis C untuk lebih membuktikan bagian yang lebih banyak termutasi, sehingga dapat dijadikan informasi untuk perlakuan *treatment* yang tepat. Berdasarkan hasil *alignment* per region, didapatkan hasil variasi genetik dimana diprediksi menunjukkan *mutation site* pada masing- masing genom, berikut merupakan hasil persentase kesamaan dan perbedaan *site* pada masing- masing genom.



Gambar 4.2 Perbandingan daerah yang terkonservasi dengan daerah yang diduga mengalami mutasi

Berdasarkan gambar 4.2, dapat dilihat warna biru pada diagram batang menunjukkan kesamaan atau *conserve* dan bagian yang berwarna merah menunjukkan persentase dalam site diduga termutasi. Pada beberapa *site* yang menunjukkan lebih tingginya sekuen yang termutasi, maka secara keseluruhan, genom virus hepatitis C mudah termutasi. Hal ini sesuai dengan penelitian dari Cuevas dkk, (2009) yang menyatakan bahwa materi genetik RNA memiliki *mutation rate* yang tinggi yang dapat memunculkan diversitas genetik yang tinggi pula dan memengaruhi tingkat evolusi. Hal ini juga akan berpengaruh pada virus yang memiliki genom RNA, yang perlu penelitian lebih untuk mengatasi masalah yang ditimbulkan seperti pemberian obat pada penderita.

4.3 Analisis *Molecular Mimicry* dengan BLAST dan Bepipred

Analisis kemiripan molekul (protein) antara *Homo sapiens* dengan poliprotein Virus Hepatitis C dilakukan dengan menggunakan fasilitas BLAST yang dapat mencari kemiripan struktur asam amino pada dua organisme tersebut. Berdasarkan hasil, terdapat kemiripan molekul antara protein yang ada di manusia dengan poliprotein virus Hepatitis C. *Molecular mimicry* ter jadi pada protein intraseluler maupun ekstraseluler. Pada penelitian ini dipilih protein mimic yang berada di ekstraseluler, karena bagian ekstraseluler yang akan dikenali terlebih dahulu oleh sel- sel imunokompeten. Jika sel imunokompeten yang telah mengenali protein virus, maka bagian yang mimic pada extrseluler manusia akan juga dikenali terlebih dahulu, terutama bagian pada permukaan sel yang akan menginisiasi kerusakan atau lisisnya sel.

Berikut merupakan hasil analisis *alignment* poliprotein Virus Hepatitis C dengan protein *Homo sapiens*

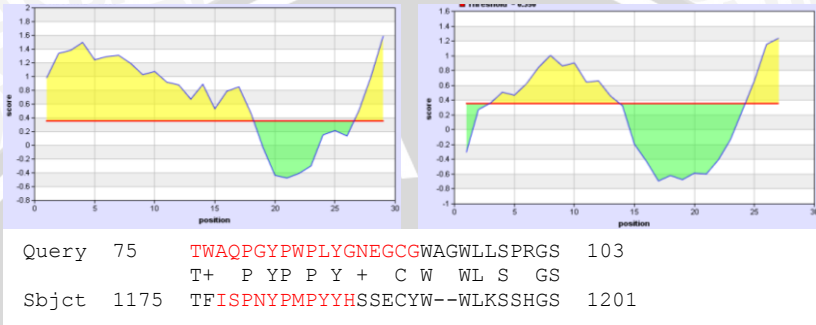
Tabel 4.3 Protein *mimic* (*Polyprotein HCV- Homo sapiens*)

Genom	Polyprotein	Mimic Protein
IA	CAPSID	intrinsic factor-B12 receptor precursor [Homo sapiens]
	NS3	G protein-coupled receptor kinase 1 [Homo sapiens]
		Insulin-like growth factor binding protein 1 [Homo sapiens]
	RNA dependen RNA Polimerase	TGF-beta type II receptor [Homo sapiens]

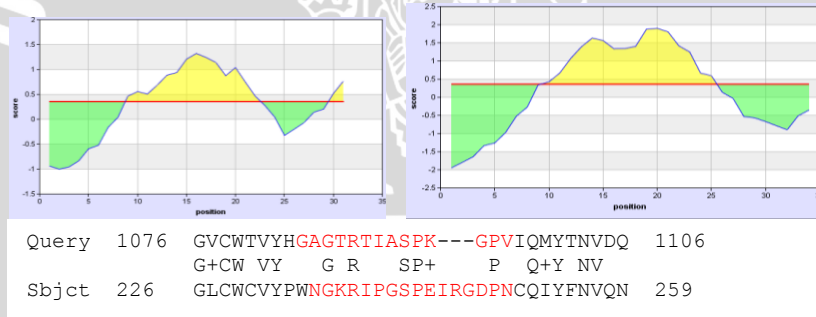
1B	DEAD Box	olfactory receptor 52E2 [Homo sapiens]
		seven transmembrane helix receptor [Homo sapiens]
	NS5A 1B	olfactory receptor 9Q2 [Homo sapiens]
2A	NS5A	olfactory receptor 9G4 [Homo sapiens]
3A	NS4B	actin, aortic smooth muscle [Homo sapiens]
	NS5	olfactory receptor 8D1
		seven transmembrane helix receptor [Homo sapiens]
4A	NS4A	TNF receptor-associated protein 1
	Peptidase S- 29/NS3	G protein-coupled receptor kinase
	RNA dependen RNA Polimerase	G protein-coupled receptor LGR5
5A	NS5A	olfactory receptor, family 5, subfamily W, member 2
	NS5A 1A	tumor necrosis factor receptor superfamily member 21
6A	CAPSID	intrinsic factor-B12 receptor precursor
	DEXDC	seven transmembrane helix receptor [Homo sapiens]
		olfactory receptor 52E2 [Homo sapiens]
	NS5A	Receptor (G protein-coupled) activity modifying protein 3 [Homo sapiens]
		cardiac sodium channel alpha subunit [Homo sapiens]
10	CAPSID	intrinsic factor-vitamin B12 receptor [Homo sapiens]
	NS2	extracellular matrix protein QBRICK [Homo sapiens]
	NS5A	olfactory receptor OR7-15 [Homo sapiens]
11	HelicC	olfactory receptor 4F5 [Homo sapiens]
	NS1	G protein-coupled receptor 123, isoform CRA_a [Homo sapiens]
	Peptidase S- 29/NS3	IGF-binding preprotein (AA -25 to 234) [Homo sapiens]

Setelah didapatkan protein yang mimic tersebut, dilakukan uji konfirmasi epitop (Lampiran 3 dan 4) dari masing- masing protein tersebut untuk mengetahui kemampuan protein-protein tersebut untuk memicu reaktifitas sel B. Berdasarkan hasil analisis ini, didapatkan yaitu *intrinsic factor B12*, *G protein-coupled receptor kinase 1*, *Insulin-like growth factor binding protein 1*, *G protein-coupled receptor LGR5*, dan *Tumor Necrosis Factor Receptor superfamily member 21*. Beberapa protein tersebut memiliki epitop

dan asam amino yang serupa antara protein *Homo sapiens* dengan Polyprotein Virus Hepatitis C. Sehingga dapat diasumsikan bahwa kedua protein tersebut *mimic* dan memiliki epitop yang serupa untuk memicu aktivasi sel B.



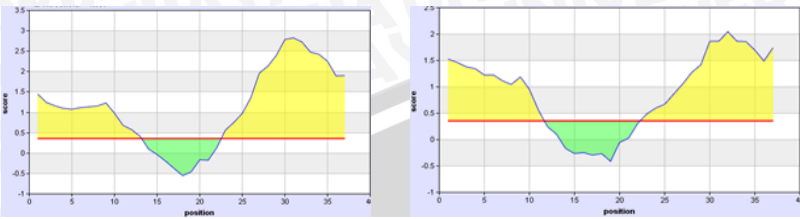
(a)



(b)

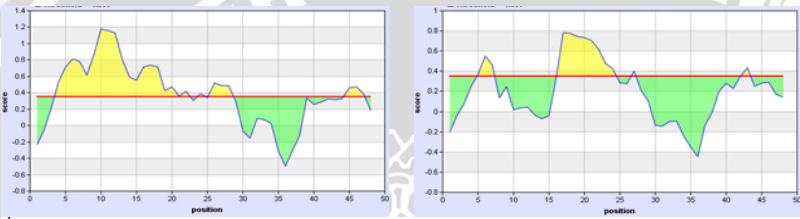


(c)



Query 2743 SDGVE----EDNRLRAFTEAMTRYSAPPGDAPQPAY 2775
 SD VE + +AL FT + Y PP P PAY
 Sbjct 854 SDDVEKQSCDS TQALVTFTSSSITYDLPPSSVPSPAY 890

(d)



tumor necrosis factor receptor superfamily member 21
 Query 2016 YKGVWRGDG--VNSTKCPGATISGHVKNGTMRIVGPKLCSNTWQGTFF 2061
 Y+V R G + KCP G +S H N ++R+ CS+ GTF
 Sbjct 54 YRHVDRATGQVLICDKCPAGTYVSEHCNTSLRV----CSSCPVGTFF 96

(e)

Gambar 4.3 Hasil analisis spitop sel B dengan Bepipred dan sekuen masing- masing protein

Keterangan: Beberapa protein memiliki kesamaan ekspresi epitope, baik pada protein *Homo sapiens* maupun pada poliprotein virus. Sumbu axis menunjukkan urutan sekuen, sumbu ordinat : skor epitope. Gambar kanan: protein *Homo sapiens* (Subject) ,kiri: poliprotein Virus Hepatitis C (query). Sekuen berwarna merah menunjukkan sekuen yang diduga memunculkan epitope.

- Letak epitop pada protein *intrinsic factor-B12 receptor precursor*
- Letak epitope pada protein *Insulin-like growth factor binding protein 1*
- Letak epitope pada protein *G protein-coupled receptor kinase 1*
- Letak epitope pada protein *G protein-coupled receptor LGR5*
- Letak epitope pada protein *tumor necrosis factor receptor superfamily member 21*

Protein *G protein coupled receptor kinase 1* dan *G protein coupled receptor LGR5* merupakan protein dalam satu famili yang tergabung di dalam G Protein Receptor. Reseptor penciuman memiliki kelas Reseptor yang disebut *G Protein Coupled Receptor* (GPCR). Protein ini mencakup protein *seven a transmembrans receptor* yang terletak pada membrane sel sebagai reseptor dan

penginduksi signal transduksi pada intrselular (Universal Protein KnowledgeBase, 2011). Berdasarkan data *Genebank* (NCBI) protein-protein tersebut masuk ke dalam cangkupan GPCR. Protein ini memiliki 662 gen terdapat pada *Olfactory organ*. Meskipun pada organ lain terdapat protein ini, tetapi berdasarkan hasil informatika, protein ini masuk ke dalam protein olfactory. Protein ini terdapat di seluruh tubuh sehingga diduga kurang dapat menimbulkan *mimicry*.

Pada hasil yang didapatkan terdapat protein *intrinsic factor-B12 receptor precursor* merupakan protein yang berfungsi untuk *uptake* vitamin B12. Reseptor ini biasa disebut dengan *cubilin receptor*. Reseptor ini terletak pada ginjal dan pada usus sebagai absorber B12 dari makanan. Jika letak reseptor ini banyak pada ginjal, diduga protein inilah yang nantinya akan memunculkan autoantibodi yang disebut Anti *Liver- Kidney Microsomal* (Anti-LKM *antibodies*). Anti LKM merupakan salah satu antibodi yang terdapat pada serum penderita autoimun Hepatitis seperti pada penelitian Zachou dkk, (2004).

Insulin Like Growth Factor Binding Protein 1 merupakan suatu protein yang dapat melakukan kompleks binding dengan *Insulin Like Growth Factor* (IGF) 1 dan 2. IGF *binding protein* dapat terletak di dalam liver (hepar) untuk mengikat IGF-1, protein ini juga berperan dalam menginisiasi *Growth Hormon* pada liver untuk memproduksi IGF-1. Hal ini penting dilakukan untuk pertumbuhan dan perkembangan femur dan otot. Protein ini banyak terdapat di hepar, tetapi karena protein ini merupakan protein sitoplasma, sehingga kurang memungkinkan sebagai protein *mimic*.

Protein *Tumor necrosis factor receptor superfamily member 21* adalah salah satu protein yang diduga *mimic* dengan protein *Homo sapiens* saat infeksi virus Hepatitis C. Protein ini sering disebut *Death Receptor- 6* (DR-6). DR- 6 banyak terekspresi di organ limfoid, hepar, jantung dan pankreas (Benschop , dkk., 2009). Protein ini dapat diduga sebagai salah satu penyebab rusaknya sel hepatosit karena protein ini *mimic* dan nantinya juga akan dikenali sebagai protein antigen oleh sel imunokompeten. Munculnya autoantibodi seperti *anti- Smooth Muscle Antibody* dan *anti nuclear antibody* dapat terjadi karena organ yang dikenali sebagai antigen

oleh sel imunokompeten adalah organ yang memiliki jaringan otot polos dan berinti tunggal.

Berdasarkan hasil dan studi pustaka yang dilakukan pada penelitian ini, polyprotein virus hepatitis C diduga *mimic* dengan protein *Homo sapiens*. Dari potensi mimicry dapat diduga pula munculnya autoimun hepatitis dapat berasal dari infeksi virus hepatitis C yang dapat memunculkan *autoantibody specific organ* (hepar) maupun *non specific organ*.

UNIVERSITAS BRAWIJAYA



BAB V PENUTUP

5.1 Kesimpulan

Polyprotein virus hepatitis C memiliki 17 *highly conserved region* dengan parameter 10 asam amino atau 30 nukleotida. Sepuluh daerah dari 17 sekuen asam amino terletak pada *functional region* dan 7 lainnya terletak pada *non functional region*. Penomoran genom pada virus hepatitis C diduga bukan berdasarkan kemiripan struktur atau komponen asam amino tetapi berdasarkan urutan waktu penemuan genom virus. Mutasi pada polyprotein virus hepatitis C terdapat hampir pada seluruh region. Mutasi banyak terjadi pada NS4 *region*, NS2 *region*, dan *envelope region*.

Infeksi virus hepatitis C diduga dapat menginduksi terjadinya *molecular mimicry* sebagai salah satu penyebab autoimun hepatitis. Protein *intrinsic factor-B12 receptor precursor*, *Tumor necrosis factor receptor superfamily member 21* merupakan protein yang diduga dapat menimbulkan reaksi *mimicry*, sedangkan protein *G protein-coupled receptor kinase 1*, *G protein-coupled receptor LGR5* dan *Insulin-like growth factor binding protein 1* diduga tidak berperan sebagai protein yang *mimic* karena persebarannya hampir di seluruh tubuh dan terletak pada sitoplasma.

5.2 Saran

Pada penelitian selanjutnya sebaiknya membandingkan antar metode yang telah dimiliki untuk *profiling polyprotein* yang lebih *advance* dan dapat mengkonfirmasi munculnya autoantibodi pada *mimicry*. Data yang digunakan dapat diuji konfirmasi dalam kondisi *in vitro* dan *in vivo*.

DAFTAR PUSTAKA

- Abbas A. K. Andrew H. Lichman ,dan Shiv Pillai. 2009. **Cellular and Molecular Immunology 6th Edition: updated Edition**. Saunders Elsevier. Philadelphia.
- Baratawidjaja, K. G. 2006. **Imunologi Dasar: Edisi Ketujuh**. Balai Penerbit FKUI. Jakarta.
- Benschop .R, Wei T., dan Na S. 2009. **Tumor Necrosis Factor Receptor Superfamily member 21**. Adv Exp Med Biol. 647:186-94.
- Blight K. J. dan Norgard E. A. 2006. HCV Replicon System. http://www.horizonbioscience.com/HCV_replicon_system. diakses tanggal 21 Juni 2010.
- Chevaliez S. dan Pawlotsky. 2006. Virus Hepatitis C Genome and Life-Cycle. http://www.horizonbioscience.com/HCV_Genome_and_Life_Cycle. diakses tanggal 21 Juni 2010.
- Cuevas J. M., P. D. Calap, M. P Calap. and R. Sanjuán. 2009. Experimental Evolution and Population Genetics of RNA Viruses. *The Open Evolution Journal*. (3): 9-16.
- Hermawan, Guntur. 2004. **Perspektif Masa Depan Imunologi-Infeksi**. Sebelas Maret University Press. Surakarta.
- ICTVdB Management. 2006. 00.026.0.03.001. Hepatitis C virus. In: ICTVdB - The Universal Virus Database, version 4. Büchen-Osmond, C. (Ed), <http://www.ncbi.nlm.nih.gov/ICTVdb/ICTVdB/> Columbia University, New York, USA. diakses tanggal 05 Desember 2010.
- Jones, G. J. Stephen V. Gordon, R. Hewinson G. ,dan H. M. Vordermeier. 2010. Screening of Predicted Secreted Antigens from Mycobacterium bovis Reveals the Immunodominance of the ESAT-6 Protein Family. *Infection and Immunity*. 78 (3): 1326–1332.
- Kresno. S.B. 2001. **Imunologi: Diagnosis dan Prosedur Laboratorium**. Balai Penerbit Fakultas Kedokteran Universitas Indonesia , Jakarta.

- Lipman. D.J., Altschul S.F., Kececioglu J.D. 1989. A tool for multiple sequence alignment. *Proc. Natl. Acad. Sci.* 86:4412-4415.
- MEGA. 2010. MEGA *Phylogenetic and Molecular Evolutionary Analysis*. <http://www.megasoftware.net/>. Diakses tanggal 05 Desember 2010.
- Meine. 2009. BioEdit Tutorial. <http://chemistry.umeche.maine.edu/CHY431/BioEdit.html>. diakses tanggal 05 Desember 2010.
- Miyaaki, H. 2009. Study Of Liver- Targeted Regulatory T Cells in Hepatitis B and C Virus in Chronically Infected Patients. *Liver International Journal*. 29(5):702-707.
- Ogden, T.H. dan Rosenberg M.S.. 2006. Multiple sequence alignment accuracy and phylogenetic inference. *Syst. Biol.* 55(2): 314-328.
- Thompson J.D. ,Higgins D.G. ,dan Gibson T.J. 1994. ClustalW—improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673-4680.
- Universal Protein KnowledgeBase. 2011. Olfactory Receptor Review. <http://www.uniprot.org/uniprot/Q8NGJ2>. Tanggal akses 06 Juni 2011.
- University of Virginia. 2007. Liver, Biliary, and Pancreatic Disorders. http://www.healthsystem.virginia.edu/uvahealth/adult_liver/home.cfm. Diakses tanggal 26 Mei 2010.
- WHO. 2010. The Hepatitis C Virus. <http://www.who.int/entity/csr/disease/hepatitis/en/>. Diakses tanggal 21 Juni 2010.
- Yang Z. dan Kumar S. 1996. Approximate methods for estimating the pattern of nucleotide substitution and the variation of substitution rates among sites. *Mol Biol Evol* 1996;13:650-659.
- Zachou, K. Rigopoulao, E. dan Dalekos G. N. 2004. Autoantibodies and Autoantigens in Autoimmune Hepatitis: Important Tools in Clinical Practice and to Study Pathogenesis of The Disease. *Journal of Autoimmune Diseases* 2004, 1(2). 17 pp.

LAMPIRAN 1 Hasil Analisis *Conserved Region* dengan *BioEdit*

BioEdit version 7.0.5.3 (10/28/05)
Conserved region search
Alignment file: Untitled1
5/10/2011 1:20:50 PM

Minimum segment length (actual for each sequence): 10
Maximum average entropy: 0.2
Maximum entropy per position: 0.2
Gaps limited to 2 per segment
Contiguous gaps limited to 1 in any segment

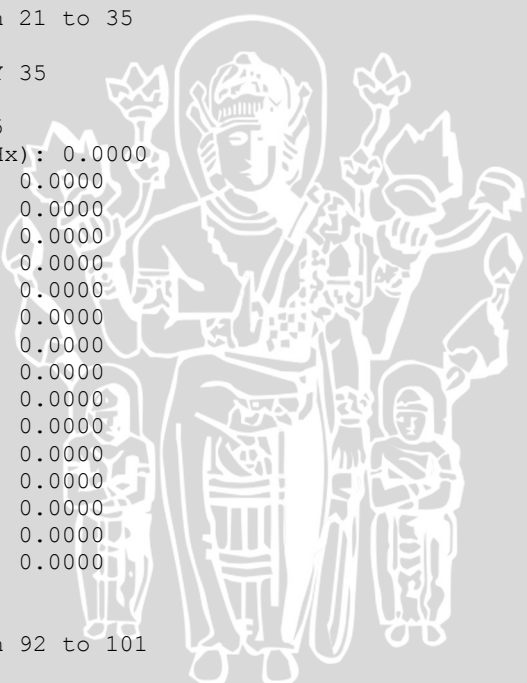
17 conserved regions found

Region 1: Position 21 to 35
Consensus:
21 DVKEFPGGGQIVGGVY 35

Segment Length: 15
Average entropy (Hx): 0.0000
Position 21 : 0.0000
Position 22 : 0.0000
Position 23 : 0.0000
Position 24 : 0.0000
Position 25 : 0.0000
Position 26 : 0.0000
Position 27 : 0.0000
Position 28 : 0.0000
Position 29 : 0.0000
Position 30 : 0.0000
Position 31 : 0.0000
Position 32 : 0.0000
Position 33 : 0.0000
Position 34 : 0.0000
Position 35 : 0.0000

Region 2: Position 92 to 101
Consensus:
92 GWAGWLLSPR 101

Segment Length: 10
Average entropy (Hx): 0.0000
Position 92 : 0.0000
Position 93 : 0.0000



Position 94 : 0.0000
Position 95 : 0.0000
Position 96 : 0.0000
Position 97 : 0.0000
Position 98 : 0.0000
Position 99 : 0.0000
Position 100 : 0.0000
Position 101 : 0.0000

Region 3: Position 504 to 517
Consensus:
504 VCGPVYCFTPSEVV 517

Segment Length: 14
Average entropy (Hx): 0.0000
Position 504 : 0.0000
Position 505 : 0.0000
Position 506 : 0.0000
Position 507 : 0.0000
Position 508 : 0.0000
Position 509 : 0.0000
Position 510 : 0.0000
Position 511 : 0.0000
Position 512 : 0.0000
Position 513 : 0.0000
Position 514 : 0.0000
Position 515 : 0.0000
Position 516 : 0.0000
Position 517 : 0.0000

Region 4: Position 588 to 597
Consensus:
588 CPTDCFRKHP 597

Segment Length: 10
Average entropy (Hx): 0.0000
Position 588 : 0.0000
Position 589 : 0.0000
Position 590 : 0.0000
Position 591 : 0.0000
Position 592 : 0.0000
Position 593 : 0.0000
Position 594 : 0.0000
Position 595 : 0.0000

Position 596 : 0.0000
Position 597 : 0.0000

Region 5: Position 1258 to 1268

Consensus:
1258 VLVLNPSVAAT 1268

Segment Length: 11
Average entropy (Hx): 0.0000
Position 1258 : 0.0000
Position 1259 : 0.0000
Position 1260 : 0.0000
Position 1261 : 0.0000
Position 1262 : 0.0000
Position 1263 : 0.0000
Position 1264 : 0.0000
Position 1265 : 0.0000
Position 1266 : 0.0000
Position 1267 : 0.0000
Position 1268 : 0.0000

Region 6: Position 1337 to 1347

Consensus:
1337 GTVLDQAETAG 1347

Segment Length: 11
Average entropy (Hx): 0.0000
Position 1337 : 0.0000
Position 1338 : 0.0000
Position 1339 : 0.0000
Position 1340 : 0.0000
Position 1341 : 0.0000
Position 1342 : 0.0000
Position 1343 : 0.0000
Position 1344 : 0.0000
Position 1345 : 0.0000
Position 1346 : 0.0000
Position 1347 : 0.0000

Region 7: Position 1393 to 1409

Consensus:
1393 KGRHLIFCHSKKCCDE 1409

Segment Length: 17
Average entropy (Hx): 0.0000
Position 1393 : 0.0000
Position 1394 : 0.0000
Position 1395 : 0.0000
Position 1396 : 0.0000
Position 1397 : 0.0000
Position 1398 : 0.0000
Position 1399 : 0.0000
Position 1400 : 0.0000
Position 1401 : 0.0000
Position 1402 : 0.0000
Position 1403 : 0.0000
Position 1404 : 0.0000
Position 1405 : 0.0000
Position 1406 : 0.0000
Position 1407 : 0.0000
Position 1408 : 0.0000
Position 1409 : 0.0000

Region 8: Position 1452 to 1462
Consensus:
1452 TGDFDSVIDCN 1462

Segment Length: 11
Average entropy (Hx): 0.0000
Position 1452 : 0.0000
Position 1453 : 0.0000
Position 1454 : 0.0000
Position 1455 : 0.0000
Position 1456 : 0.0000
Position 1457 : 0.0000
Position 1458 : 0.0000
Position 1459 : 0.0000
Position 1460 : 0.0000
Position 1461 : 0.0000
Position 1462 : 0.0000

Region 9: Position 1627 to 1636
Consensus:
1627 GPTPLLYRLG 1636

Segment Length: 10
Average entropy (Hx): 0.0000

Position 1627 : 0.0000
Position 1628 : 0.0000
Position 1629 : 0.0000
Position 1630 : 0.0000
Position 1631 : 0.0000
Position 1632 : 0.0000
Position 1633 : 0.0000
Position 1634 : 0.0000
Position 1635 : 0.0000
Position 1636 : 0.0000

Region 10: Position 1777 to 1793
Consensus:
1777 SGIQYLAGLSTLPGNPA 1793

Segment Length: 17
Average entropy (Hx): 0.0000
Position 1777 : 0.0000
Position 1778 : 0.0000
Position 1779 : 0.0000
Position 1780 : 0.0000
Position 1781 : 0.0000
Position 1782 : 0.0000
Position 1783 : 0.0000
Position 1784 : 0.0000
Position 1785 : 0.0000
Position 1786 : 0.0000
Position 1787 : 0.0000
Position 1788 : 0.0000
Position 1789 : 0.0000
Position 1790 : 0.0000
Position 1791 : 0.0000
Position 1792 : 0.0000
Position 1793 : 0.0000

Region 11: Position 1921 to 1935
Consensus:
1921 QWMNRLIAFASRGNH 1935

Segment Length: 15
Average entropy (Hx): 0.0000
Position 1921 : 0.0000
Position 1922 : 0.0000
Position 1923 : 0.0000



Position 1924 : 0.0000
Position 1925 : 0.0000
Position 1926 : 0.0000
Position 1927 : 0.0000
Position 1928 : 0.0000
Position 1929 : 0.0000
Position 1930 : 0.0000
Position 1931 : 0.0000
Position 1932 : 0.0000
Position 1933 : 0.0000
Position 1934 : 0.0000
Position 1935 : 0.0000

Region 12: Position 2401 to 2415

Consensus:

2401 SSMPPLEGEPPGDDL 2415

Segment Length: 15

Average entropy (Hx): 0.0000

Position 2401 : 0.0000
Position 2402 : 0.0000
Position 2403 : 0.0000
Position 2404 : 0.0000
Position 2405 : 0.0000
Position 2406 : 0.0000
Position 2407 : 0.0000
Position 2408 : 0.0000
Position 2409 : 0.0000
Position 2410 : 0.0000
Position 2411 : 0.0000
Position 2412 : 0.0000
Position 2413 : 0.0000
Position 2414 : 0.0000
Position 2415 : 0.0000

Region 13: Position 2591 to 2600

Consensus:

2591 TTIMAKNEVF 2600

Segment Length: 10

Average entropy (Hx): 0.0000

Position 2591 : 0.0000
Position 2592 : 0.0000
Position 2593 : 0.0000

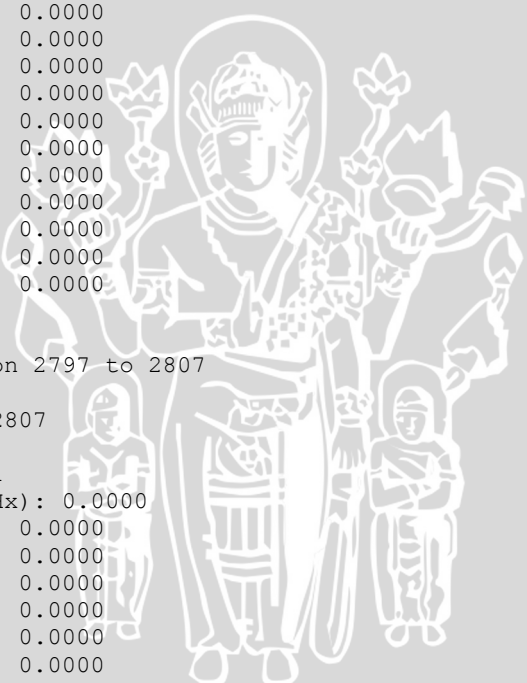
Position 2594 : 0.0000
Position 2595 : 0.0000
Position 2596 : 0.0000
Position 2597 : 0.0000
Position 2598 : 0.0000
Position 2599 : 0.0000
Position 2600 : 0.0000

Region 14: Position 2672 to 2682
Consensus:
2672 FSYDTRCFDST 2682

Segment Length: 11
Average entropy (Hx): 0.0000
Position 2672 : 0.0000
Position 2673 : 0.0000
Position 2674 : 0.0000
Position 2675 : 0.0000
Position 2676 : 0.0000
Position 2677 : 0.0000
Position 2678 : 0.0000
Position 2679 : 0.0000
Position 2680 : 0.0000
Position 2681 : 0.0000
Position 2682 : 0.0000

Region 15: Position 2797 to 2807
Consensus:
2797 AMTRYSAPPGD 2807

Segment Length: 11
Average entropy (Hx): 0.0000
Position 2797 : 0.0000
Position 2798 : 0.0000
Position 2799 : 0.0000
Position 2800 : 0.0000
Position 2801 : 0.0000
Position 2802 : 0.0000
Position 2803 : 0.0000
Position 2804 : 0.0000
Position 2805 : 0.0000
Position 2806 : 0.0000
Position 2807 : 0.0000



Region 16: Position 2819 to 2828

Consensus:

2819 TSCSSNVSA 2828

Segment Length: 10

Average entropy (Hx): 0.0000

Position 2819 : 0.0000

Position 2820 : 0.0000

Position 2821 : 0.0000

Position 2822 : 0.0000

Position 2823 : 0.0000

Position 2824 : 0.0000

Position 2825 : 0.0000

Position 2826 : 0.0000

Position 2827 : 0.0000

Position 2828 : 0.0000

Region 17: Position 2859 to 2868

Consensus:

2859 PVNSWLGNI 2868

Segment Length: 10

Average entropy (Hx): 0.0000

Position 2859 : 0.0000

Position 2860 : 0.0000

Position 2861 : 0.0000

Position 2862 : 0.0000

Position 2863 : 0.0000

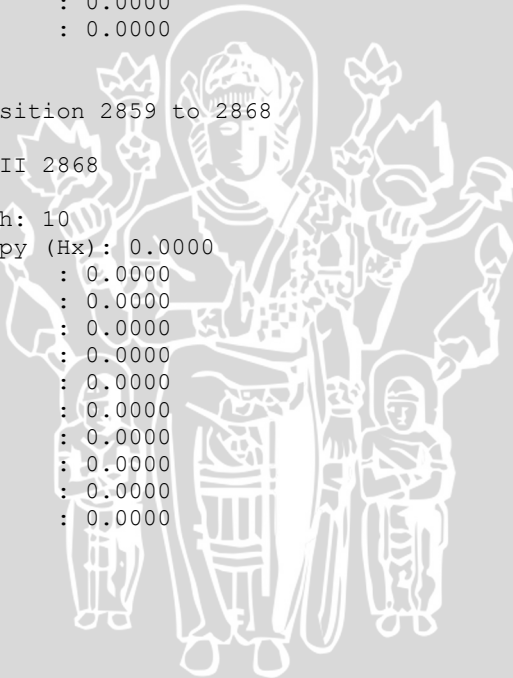
Position 2864 : 0.0000

Position 2865 : 0.0000

Position 2866 : 0.0000

Position 2867 : 0.0000

Position 2868 : 0.0000



Lampiran 2. *Alignment* region Poliprotein Virus Hepatitis C

Alignment: Capsid HCV

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
          10          20          30          40          50
gi | 329874:  STNPKPQKKN KRNTNRRPQD VKFPGGGQIV GGVYLLPRRG PRLGVRATRK
gi | 221611:  STNPKPQRKT KRNTNRRPQD VKFPGGGQIV GGVYLLPRRG PRLGVRATRK
gi | 221651:  STNPKPQRKT KRNTNRRPQD VKFPGGGQIV GGVYLLPRRG PRLGVRATRK
gi | 1183029  STLPKPQRKT KRNTIRRPQD VKFPGGGQIV GGVYVLPRRG PRLGVRATRK
gi | 2252490  STNPKPQRKT KRNTNRRPMD VKFPGGGQIV GGVYLLPRRG PRLGVRATRK
gi | 2462304  STNPKPQRKT KRNTNRRPQD VKFPGGGQIV GGVYLLPRRG PRLGVRATRK
gi | 2326455  STLPKPQRKT KRNTNRRPMD VKFPGGGQIV GGVYLLPRRG PRLGVRATRK
gi | 1183033  STLPKPQRIT KRNTNRRPQD VKFPGGGQIV GGVYVLPRRG PRLGVRATRK
gi | 1183031  STNPKPQRQT KRNTNRRPQD VKFPGGGQIV GGVYLLPRRG PRLGVRATRK

```

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
          60          70          80          90         100
gi | 329874:  TSERSQPRGR RQPIPKARRP EGRTWAQPGY PWPLYGNEGC GWAGWLLSPR
gi | 221611:  TSERSQPRGR RQPIPKARRP EGRTWAQPGY PWPLYGNEGM GWAGWLLSPR
gi | 221651:  TSERSQPRGR RQPIPKDRRS TGKSWGKPGY PWPLYGNEGL GWAGWLLSPR
gi | 1183029  TSERSQPRGR RQPIPKARRS EGRSWAQPGY PWPLYGNEGC GWAGWLLSPR
gi | 2252490  TSERSQPRGR RQPIPKARRP EGRSWAQPGY PWPLYGNEGC GWAGWLLSPR
gi | 2462304  NSERSQPRGR RQPIPKARRP TGRSWGQPGY PWPLYANEGL GWAGWLLSPR
gi | 2326455  TSERSQPRGR RQPIPKARQP QGRHWAQPGY PWPLYGSEGC GWAGWLLSPR
gi | 1183033  TSERSQPRSR RQPIPRARRT EGRSWAQPGY PWPLYGNEGC GWAGWLLSPR
gi | 1183031  TSERSQPRGR RQPIPKARRQ TGRAWGQPGY AWPLYGNEGC GWAGWLLSPR

```

```

.....|.....| .....
          110
gi | 329874:  GSRPSWGPTD PRRR
gi | 221611:  GSRPSWGPTD PRRR
gi | 221651:  GSRPSWGPND PRHR
gi | 1183029  GSRPSWGPND PRRR
gi | 2252490  GSRPSWGPND PRGR
gi | 2462304  SSRPNWGPND PRRK
gi | 2326455  GSRPHWGPND PRRR
gi | 1183033  GSRPSWGPND PRRR
gi | 1183031  GSRPTWGPND PRRR

```

Alignment: Core

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
          10          20          30          40          50
gi | 329874:  SRNLGKVIDT LTCGFADLMG YIPLVGAPLG GAARALAHGV RVLEDGVNYA
gi | 221611:  SRNLGKVIDT LTCGFADLMG YIPLVGAPLG GAARALAHGV RVLEDGVNYA
gi | 221651:  SRNVGKVIDT LTCGFADLMG YIPVVGAPLG GVARALAHGV RVLEDGVNFA
gi | 1183029  SRNLGKVIDT LTCGFADLMG YIPLVGAPVG GVARALAHGV RALEDDGINFA
gi | 2252490  SRNLGKVIDT LTCGFADLMG YIPLVGAPVG SVARALAHGV RALEDDGINYA

```



```

.....|.....|.....|.....|.....|
      110      120      130      140      150
gi | 329874: FSPRRHWTQ GCNCSIYPGH ITGHRMAWDM MMNWSPTTAL VMAQLLRIPQ
gi | 221611: FSPRRYETVQ DCNCSIYPGH VSGHRMAWDM MMNWSPTTAL VVSQLLRIPQ
gi | 221651: VSPQHWHFVQ DCNCSIYPGT ITGHRMAWDM MMNWSPTATM ILAYAMRVPE
gi | 1183029 FRPRRHQTVQ TCNCSLYPGH LSGHRMAWDM MMNWSPAVGM VVAHVLRLPQ
gi | 2252490 FRPRRHWTQ DCNCSIYTGH ITGHRMAWDM MMNWSPTTTL VLAQVMRIPT
gi | 2462304 YKPRQHHTVQ DCNCSIYSGH ITGHRMAWDM MMKWSPTTAL LMAQLLRIPQ
gi | 2326455 FQPRRHWTVQ DCNCSIYTGH VTGHKMAWDM MMNWSPTTTL VLSSILRVPE
gi | 1183033 WRHRQHWTVQ DCNCSIYPGH LTGHRMAWDM MMNWSPAMTL IVSQVLRLPQ
gi | 1183031 FRPRIHQTVQ DCNCSIYTGH VTGHRMAWDM MMNWSPTATF VVSSALRAPQ

```

```

.....|.....|.....|.....|
      160      170      180      190
gi | 329874: AILDMIAGAH WGVLAGIAYF SMVGNWAKVL VVLLLFAGVD
gi | 221611: AVVDMVAGAH WGVLAGLAYY SMVGNWAKVL IVMLLFAGVD
gi | 221651: VIIDIIGGAH WGVMFGLAYF SMQGAWAKVV VILLLAAGVD
gi | 1183029 TLFDIMAGAH WGILAGLAYY SMQGNWAKVA IIMVMFSGVD
gi | 2252490 TLVDLLSGGH WGVLVGVAYF SMQANWAKVI LVLFELFAGVD
gi | 2462304 VVIDIAGGH WGVLLAAAYF ASTANWAKVI LVLFELFAGVD
gi | 2326455 ICASVIFGGH WGILLAVAYF GMAGNWLKVL AVLFLFAGVE
gi | 1183033 TMFDLVIGAH WGVVAGVAYY SMQGNWAKVF LVLCFLFSVD
gi | 1183031 VLFDFIFAGGH WGIIGALLYY STAANWAKVI IVLLLFAGVD

```

Alignment: Helic-C

```

.....|.....|.....|.....|.....|
      10      20      30      40      50
gi | 329874: KAIPLEVIKGRHLIFCHSK KKCEDELAACL VALGINAVAY YRGLDVSVIP
gi | 221611: KAIPLEAIKGRHLIFCHSK KKCEDELAACL TGLGLNAVAY YRGLDVSVIP
gi | 221651: RAIPLSYIKGRHLIFCHSK KKCEDELAACL RGMGLNAVAY YRGLDVSVIP
gi | 1183029 KAIPIALKGRHLIFCHSK KKCEDELAACL RGMGLNAVAY YRGLDVSVIP
gi | 2252490 KAIPLELIKGRHLIFCHSK KKCEDELAACL TSLGLNAVAY YRGLDVSVIP
gi | 2462304 RAIPLVLIKGRHLIFCHSK KKCEDELAACL TSLGVNAVAY YRGLDVAVIP
gi | 2326455 KAIPLEYIKGRHLIFCHSK KKCEDELAGKL KSLGLNAVAF YRGVDVSVIP
gi | 1183033 RAIPLVGIKGRHLIFCHSK KKCEDELAACL TSLGVNAVAF YRGLDVSVIP
gi | 1183031 KAIPLEYIKGRHLIFCHSK KKCEDELAACL RTLGLNAVAF YRGVDVSVIP

```

```

.....|.....|.....|.....|.....|
      60      70      80      90      100
gi | 329874: TSGDVVVVAT DALMTGYTGD FDSVIDCN-----
gi | 221611: TSGDVVVVAT DALMTGFTGD FDSVIDCN-----
gi | 221651: TQGDVVVVAT DALMTGFTGD FDSVIDCNVA VTQVVDVDFSLD PTFTITTTQTV
gi | 1183029 TTGDVVVCAT DALMTGFTGD FDSVIDCN-----
gi | 2252490 TSGDVVVVAT DALMTGFTGD FDSVIDCNTS VIQTVD-----
gi | 2462304 ATGDVVVCST DALMTGFTGD FDSVIDCN-----
gi | 2326455 TSGDVVVVAT DALMTGYTGD FDSVIDCNVA VTQVVDVDFSLD PTFSIET---
gi | 1183033 TQGDVVVCAT DALITGYTGD FDSVIDCN-----
gi | 1183031 TSGDVVVVAT DALMTGYTGD FDSVIDCNVA VTQIVDFSLD PTFSIETTTV

```

```

.....|.....| .....|.....| .....
                110          120
gi | 329874:  -----
gi | 221611:  -----
gi | 221651:  -----
gi | 1183029: -----
gi | 2252490: -----
gi | 2462304: -----
gi | 2326455: -----
gi | 1183033: -----
gi | 1183031: PQDAVARSQR  RGRTGRGKPG  VYRY

```

Alignment: NS1

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
                10          20          30          40          50
gi | 329874:  HVTGGSAGHT  VSGFVSL LAP  GAKQNVQLIN  TNGSWHLNST  ALNCNDSLNT
gi | 221611:  HVTGGRVASS  TQSLVSWLSQ  GPSQKIQLVN  TNGSWHINRT  ALNCNDSLQT
gi | 221651:  HTVGGSTAHN  ARTLTGMFSL  GARQKIQLIN  TNGSWHINRT  ALNCNDSLHT
gi | 1183029:  YTTGGTASRH  TQAFAGLFDI  GPQQLQLVN  TNGSWHINST  ALNCNESINT
gi | 2252490:  HVSGAAVGRS  TAGLANLFSS  GSKQNLQLIN  SNGSWHINRT  ALNCNDSLNT
gi | 2462304:  HTVGGTVGGT  LKSLTSFFNP  GPQRQLQEVN  TNGSWHINRT  ALNCNDSLQT
gi | 2326455:  -----VSQT  TSGFASLLTP  GAKQNIQLIN  TNGSWHINRT  ALNCNDSLQT
gi | 1183033:  -ITGVAASG  AFTTISLFST  GAKQPLHLVN  TNGSWHINRT  ALNCNDSLNT
gi | 1183031:  --VASSVSQA  TSLVSLFSA  GARQNLQLIN  TNGSWHINRT  ALNCNDSLQT

```

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
                60          70          80          90          100
gi | 329874:  GWLAGLFYHH  KFNSSGCPER  LASCRLP LDF  DQGWGPISYA  N-GSGP-DQR
gi | 221611:  GFIAALFYAH  RFNASGCPER  MASCRPIDEF  AQGWGPITHD  M-PESS-DQR
gi | 221651:  GFLASLFYTH  SFNSSGCPER  MSACRSIEAF  RVGWGALQYE  DNVNTPEDMR
gi | 1183029:  GFIAGLFYHH  KFNSTGCPQR  LSSCKPITFF  RQGWGPLTDA  N-ITGPSDDR
gi | 2252490:  GFLASLFYTH  KFNSSGCSE  LACCKSLDSY  GQGWGPLGVA  N-ISGSSDDR
gi | 2462304:  GFIAGLMYAH  KFNSSGCPER  MSSCRPLAAF  DQGWGTISYA  T-ISGPSDDK
gi | 2326455:  GFLASLFYTH  KFNSSGCPER  MAACKPLAEF  RQGWGQITHK  N-VSGPSDDR
gi | 1183033:  GFIAGLLYHH  KFNSSGCV  MSACSP LDRF  AQGWGPLGPA  N-ISGPSSEK
gi | 1183031:  GFIASLFYRN  KF NATGCPER  LSACKTLDSF  DQGWGPITYA  N-ISGPAVEK

```

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
                110          120          130          140          150
gi | 329874:  PYCWHYPPKP  CGIVPAKSV  GPVYCFTPSP  VVVGTTDRSG  APTYSWGEND
gi | 221611:  PYCWHYAPRP  CGIVPASQVC  GPVYCFTPSP  VVVGTTDRFG  APTYSWGENE
gi | 221651:  PYCWHYAPRP  CGIVSASSVC  GPVYCFTPSP  VVVGTTDRLG  APTYTWGENE
gi | 1183029:  PYCWHYAPRP  CDIVPASSVC  GPVYCFTPSP  VVVGTTDARG  VPTYTWGENE
gi | 2252490:  PYCWHYAPRP  CGIVPASSVC  GPVYCFTPSP  VVVGTTDHVG  VPTYTWGENE
gi | 2462304:  PYCWHYPPRP  CGVVPARDVC  GPVYCFTPSP  VVVGTTDRRG  CPTYNWSNE
gi | 2326455:  PYCWHYAPRP  CEVVPARSVC  GPVYCFTPSP  VVVGTTDRRG  NPTYTWGENE
gi | 1183033:  PYCWHYAPRP  CDTVPAQSV  GPVYCFTPSP  VVVGATDKRG  APTYTWGENE
gi | 1183031:  PYCWHYPPRP  CEVVSALNVC  GPVYCFTPSP  VVLGTTDRRG  NPTYTWGENE

```


	160	170	180	190	200
gi 329874:	TDVFLVNNTR	PPLGNWFGCT	WMNSTGFTKV	CGAPPCVIGG	AGNNT-----
gi 221611:	TDVLLLSNTR	PPQGNWFGCT	WMNSTGFTKT	CGGPPCNIGG	VGNNT-----
gi 221651:	TDVFLLNSTR	PPQGSWFGCT	WMNSTGYTKT	CGAPPCCRIA	DFNAS-----MD
gi 1183029	KDVFLLSQR	PPSGRWFGCS	WMNSTGFLLT	CGAPPCCNIY	GEGNPHNESD
gi 2252490	TDVFLLNSTR	PPHGAWFGCV	WMNSTGFTKT	CGAPPCEVN-	TNNGT-----
gi 2462304	TDILLLNIR	PPAGNWFNGCT	WMNSTGFVKN	CGAPPCLNGP	TGNSN-----
gi 2326455	TDVFMLLESLR	PPTGGWFGCT	WMNSTGFTKT	CGAPPCCQIV	GNYNSSANE-
gi 1183033	SDVFLLSAR	PPTFPWFNGCT	WMNGSGYVKT	CGAPPCHYIG	GREGKSNNS-
gi 1183031	TDVFMSSLR	PPAGGWYNGCT	WMNSTGFVKT	CGAPP CNIRP	NPEENRTET-

	210	220	230	240	250
gi 329874:	LHCPTDCFRK	HPDATYSRCG	SGPWITPRCL	VDYPYRLWHY	PCTINYTIFK
gi 221611:	LVCPTDCFRK	HPEATYTKCG	SGPWLTPRCM	VDYPYRLWHY	PCTVNFVTFK
gi 221651:	LLCPTDCFRK	HPDTYIKCG	SGPWLTPRCL	IDYPYRLWHY	PCTVNYTIFK
gi 1183029	LFCPTDCFRK	HPETYSRCG	AGPWLTPRCM	VDYPYRLWHY	PCTVDFRLF
gi 2252490	WHCPTDCFRK	HPETTYAKCG	SGPWITPRCL	IDYPYRLWHF	PCTANFSVFN
gi 2462304	LKCPTDCFRK	HPDATYTRCG	SGPWLTPRCL	VHYPYRLWHY	PCTVNYTIFK
gi 2326455	LCCPTDCFRK	HPEATYQRCG	SGPWVTPRCL	VDYAYRLWHY	PCTVNYTLHK
gi 1183033	LVCPTDCFRK	HPDATYNRG	AGPWLTPRCL	VDYPYRLWHY	PCTVNYTIFK
gi 1183031	LRCPTDCFRK	HPGATYAKCG	SGPWLTPRCL	VDYPYRLWHY	PCTVNYTLHK

	260	270	280	290	300
gi 329874:	IRMYVGGVEH	RLEAACNWTR	GERCDLEDRD	RSELSPLLLT	TTQWQVLPCS
gi 221611:	VRMYVGGVEH	RLNAACNWTR	GERCDLEDRD	RSELSPLLLS	TTEWQILPCS
gi 221651:	IRMYVGGVEH	RLTAACNFTR	GDRCNLEDRD	RSQLSPLLLS	TTEWAILPCT
gi 1183029	IRMFVGGFEH	RFTAACNWTR	GERCDIEDRD	RSEQHPLLHS	TELAAILPCS
gi 2252490	IRTFVGGIEH	RMQAACNWTR	GEVCGLEHRD	RVELSPLLLT	TTAWQILPCS
gi 2462304	VRMFVGGLEH	RLEAACNWTY	GERCDLEDRD	RAELSPLLHT	TTQWAILPCS
gi 2326455	VRMFVGGTEH	RFDVACNWTR	GERCELHDRN	RIEMSPLLFS	TTQLSILPCS
gi 1183033	VRMFVGGLEH	RFNAACNWTR	GERCNLEDRD	RSEMYPLLHS	TTEQAILPCS
gi 1183031	VRMYIAGSEH	RFTAACNWTR	GERCDLADRD	RIEMSPLLFS	TTELAAILPCS

	310	320	330	340	350
gi 329874:	FTTLPALSTG	LIHLHQNIVD	VQYLYGVGSS	IASWAIKWEY	VVLLFLLLAD
gi 221611:	FTTLPALSTG	LIHLHRNIVD	VQYLYGIGSA	VVSFAIKWEY	ILLFLLLAD
gi 221651:	YSDLPALSTG	LLHLHQNIVD	VQFMYGLSPA	LTKYIVRWEW	VVLLFLLLAD
gi 1183029	FTTPMPALSTG	LIHLHQNIVD	VQYLYGVGSS	MVGWALKWEF	VILVFLLLAD
gi 2252490	FTTLPALSTG	LIHLHQNIVD	VQYLYGVGSA	VVSWALKWEF	VVLAFLLLAD
gi 2462304	FTTPPALSTG	LIHLHQNIVD	TQYLYGLSS	IVSWAVKWEY	IMLVFLLLAD
gi 2326455	FSTMPALSTG	LIHLHQNIVD	VQYLYGVSTN	VTSVWVWWEY	IVLMFLVLAD
gi 1183033	FVPIPALSTG	LIHLHQNIVD	VQYLYGISSG	LVGWAIKWEF	VILIFLLLAD
gi 1183031	FTTMPALSTG	LIHLHQNVVD	VQYLYGLSTS	IVNWAIKWEY	VVLLFLVLAD

gi | 329874: A
 gi | 221611: A
 gi | 221651: A
 gi | 1183029 A
 gi | 2252490 A
 gi | 2462304 A
 gi | 2326455 A
 gi | 1183033 A
 gi | 1183031 S

Alignment: NS4A

```

  .....|.....| .....|.....| .....|.....| .....|.....|
             10         20         30         40         50
gi | 329874: STWVLVGGVLAALAAAYCLSTGCVVIVGRVVLGSKPAIIPDREVLVREFDDE
gi | 221611: STWVLVGGVLAALAAAYCLTTGSVVIVGRILSGRPVAVIPDREVLVQEFDE
gi | 221651: -----CVCIIIGRLHVNRQAVVAPDKEVLVYEAFFDE
gi | 1183029 STWVLVGGVLAALAAAYCLSVGCVVIVGHIIELEGKPAVLPDKEVLVYQQYDE
gi | 2252490 -----PAVIPDREVLVYQQFDE
gi | 2462304 STWVLVGGVVAALAAAYCLTVGSVAIVGRILSGRPVAVIPDREVLVYQQFDE
gi | 2326455 STWVLVGGVLAALAAAYCLSVGCVVICGRITLTGKPAVVPDREILVYQQFDE
gi | 1183033 SAWVLVGGVMAALTAAYCLSVGSVVIVGHLVVGKPAVLPDKEVLVYQQYDE
gi | 1183031 STWVAVGGVLAATAAYCLTVGSVICGRITTSRPAVVPDREVMVYQQYDE

```

....

gi | 329874: MEEC
 gi | 221611: MEEC
 gi | 221651: MEEC
 gi | 1183029 MEEC
 gi | 2252490 MEEC
 gi | 2462304 MEEC
 gi | 2326455 MEEC
 gi | 1183033 MEEC
 gi | 1183031 MEEC

Alignment: NS2

```

  .....|.....| .....|.....| .....|.....| .....|.....|
             10         20         30         40         50
gi | 329874: DTEVAASC GG VVLVGLMALT LSPYYKRYIS WCLWWLQYFL TRVEAQLHVW
gi | 221611: DREMAASCGG AVFVGLVLLT LSPYYKVF LA RLWWLQYFI TRAE AHLQVW
gi | 221651: --SVHGQIGA ALLVLTITLFT LTPGYKTLIS RFLWWLQYFL TLAEAMVQEW
gi | 1183029 -GEDSATLGA GVLVLF GFFT LSPWYKHWIG RLMWNNQYTI CRCESALHVW
gi | 2252490 DQEVAGSLGG AIVVMLTILT LSPHYKWLVA RGLWWIQYFI ARTEAVLHVY
gi | 2462304 DSSDGGTVGC LVLIVLTIFT LTPGYKWKVV LVMWNLQYFI ARVEAIIHVW
gi | 2326455 DNEQAASLGA VGLLVITIFS ITPMYKLLN CFIWNNQYFL ARAEAMVHEW
gi | 1183033 -GTDSTLGA GVLSLFALFT LSPWYKHWIA RLWNNQYTI ARCEAALQIW
gi | 1183031 DGVQAGSIGA AVIAALTIFS LTPAYKTLIA HFLWWTQYFI AHIEAKLHVW

```


	60	70	80	90	100
gi 329874:	IPPLNVRGGR	DAVILLMCAV	HPTLVFDITK	LLLAVFGPLW	ILQASLLKVP
gi 221611:	VPPLNVRGGR	DAIILLTCAV	HPELIFDITK	LLLAILGLPM	VLQAGITRVP
gi 221651:	APPMQVRGGR	DGIIWAVAIF	CPGVVFDITK	WLLAVLGPAY	LLKGALTRVP
gi 1183029	VPPLLARGSR	DGVILLTSL	YPSLIFDITK	LLMAVLGPLY	LIQATITTTT
gi 2252490	IPSFNVRGPR	DSVIVLAVLV	CPDLVFDITK	YLLAILGLPH	ILQASLLRIP
gi 2462304	VPPLQVKGGR	DAVIMLTCLF	HPALGFEITK	ILFGILGPLY	LLQHSLLTKVP
gi 2326455	VPDLRVRGGR	DSIILLTCLL	HPQLGFEVTK	ILLAVLAPLY	ILQYSLLKVP
gi 1183033	VPPLLARGAR	DGIILLAGLF	YPALVFDITK	LLLAILGPLY	ILQASLVRVP
gi 1183031	VPFLRVRGGR	DAIILLTCVF	HPSLGFEVTK	ILLALIGPLY	LLHASLRLVP

	110	120	130	140	150
gi 329874:	YFVRVQGLLR	FCALARKMIG	GHYVQMVIIK	LGALTGTYYV	NHLTPLRDWA
gi 221611:	YFVRAQGLIR	ACMLRVKVG	GHYVQMAFMK	LAALTGTYYV	DHLTPLRDWA
gi 221651:	YFVRAHALLR	MCTMVRHLAG	GRYVQMVLLA	LGRWTGTYYI	DHLTPMSDWA
gi 1183029	YFVRAHVLR	LCMLVRSVIG	GKYFQMIILS	IGRWFNTYYI	DHLAPMQHWA
gi 2252490	YFVRAQALVK	ICSLLRGVVY	GKYFQMVVLK	SRGLTGTYYI	DHLTPMSDWP
gi 2462304	YFLRARALLR	LCLLAKHLVY	GKYVQAALLH	LGRLTGTYYI	DHLAPMKDWA
gi 2326455	YFVRAHILLR	ACLLVRRLAG	GKYVQACLRL	LGAWTGTFFVY	DHLAPLSDWA
gi 1183033	YFVRAHAVVR	LCILVRNITG	GKYVQMVLLA	LARGFNTYYI	DHLSPMTDWA
gi 1183031	YYVRAHALIR	ICALVQNVAG	GKYVQAAILR	AGSWTGTYYI	DHLVPLRTWA

	160	170	180	190
gi 329874:	HNGLRDLAVA	VEPVVFSQME	TKLITWGADT	AACGDIINGL PVSAR
gi 221611:	HAGLRDLAVA	VEPVVFSQME	TKLITWGADT	AACGDIISGL PVSAR
gi 221651:	ANGLRDLAVA	VEPIIFSPME	KKVIVWGAET	AACGDILHGL PVSAR
gi 1183029	AAGLKDLAVA	TEPVIFSPME	IKVITWGADT	AACGDILCGL PVSAR
gi 2252490	PYGLRDLAVA	LEPVVFTPME	KKVIVWGAET	AACGDIIRGL PVSAR
gi 2462304	ASGLRELTVA	TEPVIFSAME	TKVITWGADT	AACGNILAVL PVSAR
gi 2326455	SDGLRDLAVA	VEPVIFSPME	KKIITWGADT	AACGDILSGL PVSAR
gi 1183033	AEGLKDLAVA	VEPVIFSPME	VKVITWGADT	TACGDILCGL PVSAR
gi 1183031	SDGLRDLAVA	VEPVVFSQME	KKVITWGADT	AACGDILAGL PVSAR

Alignment: NS4b

```
.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
          10          20          30          40          50
gi|329874: FKQKALGLLQ TASRQAEVIA PAVQTNWQKL ETFWAKHMWN FISGIQYLAG
gi|221611: FKQKALGLLQ TATKQAEAAA PVVESKWRAL EVFWAKHMWN FISGIQYLAG
gi|221651: -KSKIQGLLQ QASKQAQDIQ PAVQASWPKV EQFWAKHMWN FISGIQYLAG
gi|1183029: FKEKILGLLQ RATQQQAVIE PIVTTNWQKL EAFWHKHMWN FVSGIQYLAG
gi|2252490: FKQKALGLLN FAGKQAQEAAT PVIQSNFAKL EQFWANDMWN FISGIQYLAG
gi|2462304: FKEKVLGLIG TAGQKAETLK PAATSMWSKA EQFWAKHMWN FVSGIQYLAG
gi|2326455: FRQKVLGLLQ ASAKQAEELK PAVHSAWPRV EDEFWRKHMWN FVSGIQYLAG
gi|1183033: FKEKVIQLLQ QADQKAADIK PIATPYWQKL ETFWSKHMWN FVSGIQYLAG
gi|1183031: FKQNVLGLIQ VTTKQAEELK PAVHSAWPKL EQFWYKHMWN FISGIQYLAG
```

```
.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
          60          70          80          90         100
gi|329874: LSTLPGNPAI ASLMAFTAASV TSPLTTSQTL LFNILGGWVA AQLAAPGAAT
gi|221611: LSTLPGNPAI ASLMAFTASI TSPLTTQNTL LFNILGGWVA AQLAPPSAAS
gi|221651: LSTLPGNPAV ASMMAFSAAL TSPLSTSTI LLNILGGWLA SQIAPPAGAT
gi|1183029: LSTLPGNPAV ASLMAFTASV TSPLTTNQT MFFNILGGWVA THLAGPQSSS
gi|2252490: LSTLPGNPAI ASLMSFTAASV TSPLTTQQT LFNILGGWVA SQIRSDAST
gi|2462304: LSTLPGNPAV ATLMSFTAASV TSPLTTHQT LFNILGGWVA SQIAPPATAAT
gi|2326455: LSTLPGNPAV ASLMSFTASL TSPLRTSQTL LLNILGGWIA AQAAPPAST
gi|1183033: LSTLPGNPAI ASLMAFTASV TSPLTTNQT LFNIMGGWVA SNLAPPAST
gi|1183031: LSTLPGNPAV AALMSFSASL TSPLTTAQT LLNVLGGWVA SQLATVPAT
```

```
.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
          110         120         130         140         150
gi|329874: AFVVGAGLAGA AIGSVGLGKV LIDILAGYGA GVAGALVAFK IMSGEVPSTE
gi|221611: AFVVGAGIAGA AVGSIGLGKV LVDILAGYGA GVAGALVAFK VMMSGEMPSTE
gi|221651: GFVVSGLVGA AVGSIGLGKV LVDILAGYGA GISGALVAFK IMSGEKPSME
gi|1183029: AFVVSGLAGA AIGGIGLGRV LLDILAGYGA GVSGALVAFK IMGGECPSTAE
gi|2252490: AFVVSGLAGA AVGSVGLGKI LVDILPGYGA GVRGAVVTFK IMSGEMPSTE
gi|2462304: AFVVSGMAGA AVGNIGLGRV LIDILAGYGT GVAGALVAFK IMCGERPTAE
gi|2326455: AFVVSGLAGA AVGSIRLGRV LVDVLGYGA GVSALVAFK IMSGECPSTE
gi|1183033: AFVVSGLAGA AVGSIGLGKV LLDILAGYGA GVAGALVAFK IMGGEMPSTE
gi|1183031: AFVVSGLAGA AIGSIGLGKV IVDILAGYGA GVSALVAFK IMSGETPSVE
```

```
.....|.....| .....|.....| .....|.....| .....|.....| .....
          160         170         180         190
gi|329874: DLVNLLPAIL SPGALVVGVI CAAILRRHVG PGEGAVQWMN RLIA
gi|221611: DLVNLLPAIL SPGALVVGVI CAAILRRHVG PGEGAVQWMN RLIA
gi|221651: DVVNLLPGIL SPGALVVGVI CAAILRRHVG PGEGAVQWMN RLIA
gi|1183029: DMVNLLPAIL SPGALVVGVI CAAILRRHVG PGEGAVQWMN RLIA
gi|2252490: DLVNLLPAIL SPGALVVEVV CPAILRRHVG PGEGAVQWMN RLIA
gi|2462304: ELVNLLPSIL CPGALVVGVI CAAVLRRHIG PGEGAVQWMN RLIA
gi|2326455: DMVNLLPALL SPGALVGVV CAAILRRHVG PAEGANQWMN RLIA
gi|1183033: DMVNLLPAIL SPGALVVGVI CAAILRRHVG PGEGAVQWMN RLIA
gi|1183031: DMVNLLPALL SPGALVVGVI RAAILRRHVG PSEGAAQWMN RLIA
```

Alignment: NS5A

```
.....|.....| .....|.....| ...
                10         20
gi|221651:   GSWLRDVWDW VCTILTDFKN WL-
gi|1183029  --WLRTIWDW VCSVLADFKA WLS
gi|2252490  -SWLWEVWDW VLVHVSDFKT CLK
gi|2462304  GTWLRAIWDW VCTALTDFKA WLQ
gi|2326455  -SWLRDVWDW VCTVLSDFKV WLQ
gi|1183033  GNWLYDIWNW VCTVLADFKL WL-
gi|1183031  STWLRDIWDW VCTVLSDFRV WLK
gi|329874:   GSWLRDIWDW ICEVLSDFKT WLK
gi|221611:   GSWLKDVWDW ICTVLSDFKT WLQ
```

Alignment: NS5a_1A

```
.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
                10         20         30         40         50
gi|329874:   IPFVSCQRGY KGVWRVDGIM HTRCHCGAEI TGHVKNGTMR IVGPRTCRNM
gi|221611:   LPFLSCQRGY KGVWRGDGIM QTTCPCGAQI TGHVKNGTMR IVGPKTCSNT
gi|221651:   LPFISCQKGY KGVWAGTGIM TTRCPCGANI SGNVRLGSMR ITGPKTCMNI
gi|1183029  LPFISCQKGY KGVWRGDGVM STRCPCGAAI TGHVKNGTMR LAGPRTCANM
gi|2252490  IPLLWSPRGY KGEWRGDGVM HTTCCPGADL AGHIKNGSMR ITGPKTCSNT
gi|2462304  VPFVSCQKGY KGVWRGDGVM STKCPCGATI SGHVKNGTMR IVGPKLCSNT
gi|2326455  IPFLSCQAGY RGVWAGDGVC HTTCTCGAVI AGHVKNGTMR ITGPKTCSNT
gi|1183033  IPFLSCQKGY RGTWRGDGVV STRCPCGALL SGHVKNGTMR LVGPRWCANT
gi|1183031  VPFVSCQRGY RGTWRGDGIC NTTCPCGASI AGHVKNGTMR IVGPRTCSNV
```

```
.....|.....| ..
                60
gi|329874:   WSGTFPINAY TT
gi|221611:   WHGTFPINAY TT
gi|221651:   WQGTFPINCY TE
gi|1183029  WHGTFPINEY TT
gi|2252490  WHGTFPINAY TT
gi|2462304  WQGTFPINAT TT
gi|2326455  WHGTFPINAT TT
gi|1183033  WHGTFPINGY TT
gi|1183031  WNGTFPINAT TT
```

Alignment: NS5A_1B

```
.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
                10         20         30         40         50
gi|329874:   GPCTPLPAPN YTFALWRVSA EEYVEIRQVG DFHYVTGMTT DNLKPCQVVP
gi|221611:   GPCTPSPAPN YSRALWRVAA EEYVEVTRVG DFHYVTGMTT DNVKPCQVVP
gi|221651:   GQCVPKPAPN FKIAIWRVAA SEYAEVTQHG SYHYITGLTT DNLKVPQQLP
gi|1183029  GPSTPCPSPN YTRALWRVAA NSYVEVRRVG DFHYITGATE DELKPCQVVP
gi|2252490  GSPVPIPAPN YKFALWRVSA EDYVEVRRVG DFHYVTGVTQ DNIKFPQVVP
gi|2462304  GPSVPAPAPN YKFALWRVGA ADYAEVRRVG DYHYITGVTQ DNLKPCQVVP
gi|2326455  GPSTPRPAPN YQRALWRVSA EDYVEVRLG DCHYVVGVTA EGLKPCQVVP
gi|1183033  GPSTPAPSVA YSRALWRVAS DSYVEVRKVG DFHYVTGTTD DGLKPCQVVP
gi|1183031  GPSIPIPAPN YKKALWRVSA TEYVEVRRVG DSHYITGVTA ENTKPCQVVP
```

```

.....|.....|.....|.....|.....|
      60      70      80      90     100
gi|329874:  SPEFFTELDG VRLHRFAPPC KPLLREEVSF RVGLHEYVPG SQLPCEPEPD
gi|221611:  APEFFTEVDG VRLHRYAPVC KPLLREEVVF QVGLNQYLVG SQLPCEPEPD
gi|221651:  SPEFFSWVDG VQIHRFAPIP KPFFRDEVSF CVGLNSFVVG SQLPCDEPEPD
gi|1183029:  AAEFFTEVDG VRLHRYAPPC KPLLRRDDITF MVGLHSYITG SQLPCEPEPD
gi|2252490:  APELFTEVDG IRIHRHAPKC KPLLRRDEVSF SVGLNSFVVG SQLPCEPEPD
gi|2462304:  SPEFFTELDG VRIHRFAPPC NPLLREEVTF SVGLHSYVVG SQLPCEPEPD
gi|2326455:  APEFFTEVDG VRIHRYAPPC KPLLRRDEVTF SVGLSNYAVG SQLPCDEPEPD
gi|1183033:  LPEFFTELDG VRLHRYAPVC RPLLRRDDVTF TVGLNSYVIG SQLPCEPEPD
gi|1183031:  APEFFTEVDG VRLHRYAPEC KPLLRRDEVTF TVGLSTYVVG SQLPCEPEPD

```

```

.
gi|329874:  V
gi|221611:  V
gi|221651:  -
gi|1183029:  V
gi|2252490:  V
gi|2462304:  V
gi|2326455:  V
gi|1183033:  V
gi|1183031:  V

```

Alignment: Peptidase s29

```

.....|.....|.....|.....|.....|
      10     20     30     40     50
gi|329874:  EGEVQIVSTA AQTFLATCIN GVCWTVYHGA GTRTIASPKG PVIQMYTNVD
gi|221611:  -GEVQVLSTA TQSFLATCVN GVCWTVYHGA GSKTLAGPKG PITQMYTNVD
gi|221651:  AGEIQVLSTV TQSFLGTTIS GVLWTVYHGA GNKTLAGSRG PVTQMYSSAE
gi|1183029:  TGEVQVLSTA TQTFLGTTVG GVIWTVYHGA GSRTLAKAH PALQMYTNVD
gi|2252490:  -GEVQVLSTA TQSFLGTAVN GVMWTVYHGA GAKTISGPKG PVNQMYTNVD
gi|2462304:  EGEVQFLSTA TQTFLGICIN GVMWTLFHGA GSKTLAGPKG PVVQMYTNVD
gi|2326455:  EGEVQVSTD TQSFLVATSIN GVMWTVYHGP GFKTLAGPKG PVCQMYTNVD
gi|1183033:  TGEIQVLSTS TQTFLGTSVG GVMWTVYHGA GSRTLAKNKR PALQMYTNVD
gi|1183031:  AGEIQILSTA TQTFLATCVN GVCWTVYHGA GSKTLAGPRG PVCQMYTNVD

```

```

.....|.....|.....|.....|.....|
      60      70      80      90     100
gi|329874:  QDLVGWPAPQ GSRSLTPCTC GSSDLYLVTR HADVIPVRRR GDSRGSLLSP
gi|221611:  QDLVGWPAPP GARSMTPTCT GSSDLYLVTR HADVVPVRRR GDSRGSLLSP
gi|221651:  GDLVGWPSPP GTKSLEPCTC GAVDLYLVTR NADVIPARRR GDKRGALLSP
gi|1183029:  QDLVGWPAPP GAKSLEPCAC GSSDLYLVTR DADVIPARRR GDSTASLLSP
gi|2252490:  QDLVGWPAPP GVRSLAPCTC GSADLYLVTR HADVIPVRRR GDTRGALLSP
gi|2462304:  KDLVGWPSPP GKGSLTRCTC GSADLYLVTR HADVIPARRR GDTRASLLSP
gi|2326455:  LDLVGWPSPP GARSLTPCNC GSSDLYLVTR EADVIPARRR GDSRAALLSP
gi|1183033:  QDLVGWPSPP GAKSLVPCTC GSADLYLITR DADVLPARRR GDSTASLLSP
gi|1183031:  QDMVGWPAPA GTRSYPCTC GASDLYLITR QADVIPARRR GDNRAGLISP

```



```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
          110          120          130          140
gi | 329874:  RPISYLGKSS  GGPLLCPAGH  AVGIFRAA VC  TRGVAKAVDF  IPVENLETT
gi | 221611:  RPISYLGKSS  GGPLLCPSGH  VVGIFRAA VC  TRGVAKAVDF  IPVESMETT
gi | 221651:  RPLSTLKGSS  GGPVLCPRGH  AVGVFRAA VC  SRGVAKSIDF  IPVETLID--
gi | 1183029  RPLACLKGSS  GGPVMCPSGH  VAGIFRAA VC  TRGVAKSLQF  IPVETLSTQ
gi | 2252490  RPISILKGSS  GGPLLCPMGH  RAGIFRAA VC  TRGVAKAVDF  VPVESLETT
gi | 2462304  RPISYLGKSS  GGPIMCPSGH  VVGVFRAA VC  TRGVAKALEF  VPEVENLETT
gi | 2326455  RPISTLKGSS  GGPIMCPSGH  VVGLFRAA VC  TRGVAKSLDF  IPVENMETT
gi | 1183033  RPLACLKGSS  GGPIMCPSGH  VAGIFRAA VC  TRGVAKALQF  IPVESLSAQ
gi | 1183031  RPISTLKGSS  GGPLLCPSGH  VVGLFRAA VC  TRGVAKALDF  VPCEAMDAT

```

Alignment: RNA Dependent RNA Polymerase

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
          10          20          30          40          50
gi | 329874:  ----- --HINSVWK  DLLEDNVTP I  DTTIMAKNE V  FCVQPEKGG R
gi | 221611:  ----- -VNHIRSVWE  DLLEDTETP I  DTTIMAKNE V  FCVQPEKGG R
gi | 221651:  ----- -VNHIKSVWK  DLLEDTQT P I  PTTIMAKNE V  FCVDPKGGK
gi | 1183029  ----- --NQIRSVWE  DLLEDTTTP I  PTTIMAKNE V  FCVDPKGG R
gi | 2252490  ----- --NHISSVWK  DLLDDNNT P I  PTTIMAKNE V  FAVNPAKGG R
gi | 2462304  AKEVRS LDKK  ALKHIEGVW Q  DLLDDSDTP L  PTTIMAKNE V  FAVEPSKGG K
gi | 2326455  ----- -VDHIPSVWE  GLEEDSDTP I  PTTIMAKNE V  FCVDPKGG R
gi | 1183033  ----- --NHINSVWE  DLLEDNTTP I  PTTIMAKNE V  FAVAPHKGG R
gi | 1183031  ----- --NHINSVWE  DLLEDNVTP I  PTTIMAKNE V  FCVDVSKGG R

```

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
          60          70          80          90          100
gi | 329874:  KPARLIVFPD  LGVRVCEKMA  LYDVVTKLPL  AVMGSSYGFQ  YSPGQRVEFL
gi | 221611:  KPARLIVFPD  LGVRVCEKMA  LYDVVSTLPLQ  AVMGPSYGFQ  YSPGQRVEFL
gi | 221651:  KAARLIVYPD  LGVRVCEKMA  LYDITQKLPQ  AVMGASYGFQ  YSPAQRVEFL
gi | 1183029  KPARLIVYPD  LGVRVCEKRA  LYDVIQKLSI  ETMGPAYGFQ  YSPQQRVERL
gi | 2252490  KPARLIVYPD  LGSRVCEKRA  LHDVIKKTAL  AVMGAA YGFQ  YSPAQRVEFL
gi | 2462304  KPARLIVYPD  LGVRVCEKRA  LYDVAQKLP T  ALMGPSYGFQ  YSPAQRVDFFL
gi | 2326455  KPARLIVYPD  LGVRVCEKMA  LYDVTQKLPQ  AVMGPAYGFQ  YSPNQRVEYL
gi | 1183033  KPARLIVYPD  LGVRICEKRA  LYDVIQKLP S  AIMGSAYGFQ  YSPKQRVEYL
gi | 1183031  KPARLIVYPD  LSVRVCEKRA  LYDVTRKLPV  AVMGAA YGFQ  YSPSQRVEYL

```

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
          110          120          130          140          150
gi | 329874:  VQAWKSKKTP  MGFSYDTRCF  DSTVTESDIR  TEEAIYQCCD  LDPQARVAIK
gi | 221611:  VNTWKS KCP  MGFSYDTRCF  DSTVTENDIR  TEESIYQCCD  LAPEARQAIR
gi | 221651:  LKAWAEKKDP  MGFSYDTRCF  DSTVTERDIR  TEESIYRACS  LPPEAHTAIH
gi | 1183029  LKMWT SKKTP  LGFSYDTRCF  DSTVTEQDIR  VEEEIYQCCN  LEPEARQVIS
gi | 2252490  LTAWKSKNDP  MGFSYDTRCF  DSTVTEKDIR  VEEEVYQCCD  LEPEARQVIT
gi | 2462304  LKAWKSKKIP  MAFSYDTRCF  DSTITEHDIM  TEESIYQCCD  LQPEARRAIR
gi | 2326455  LKMWR SKKVP  MGFSYDTRCF  DSTVTERDIR  TENDIYQSCQ  LDPVARRVVS
gi | 1183033  LKMWN SKKTP  LGFSYDTRCF  DSTVTEQDIR  VEEESIYQACD  LKDEARRVIT
gi | 1183031  LKIWR SKKTP  MGFSYDTRCF  DSTVTERDIR  TEESIYQCE  LDPVARKAIS

```


	160	170	180	190	200
gi 329874:	SLTERLYVGG	PLTNSRGENC	GYRRCRASGV	LTTSCGNLTL	CYIKARAACR
gi 221611:	SLTERLYVGG	PLTNSKGQNC	GYRRCRASGV	LTTSCGNLTL	CYLKATAACR
gi 221651:	SLTERLYVGG	PMFNKGGQTC	GYRRCRASGV	LPTSMGNLTL	CYVKALAACK
gi 1183029	SLTERLYCGG	PMFNKGAQC	GYRRCRASGV	LPTSFNGTIT	CYIKATAAAK
gi 2252490	ALTDRLYVGG	PMHNSKGDLC	GYRRCRATGV	YTTSFGNLT	CYLKATAAIR
gi 2462304	SLTQRLYCGG	PMYNSKGGQC	GYRRCRASGV	FTTSMGNLTL	CYIKALASCR
gi 2326455	SLTERLYVGG	PMANSKGGSC	GYRRCRASGV	LPTSMGNLTL	CYLKAQAACR
gi 1183033	SLTERLYCGG	PMFNKGGQHC	GYRRCRASGV	LPTSFNGTVT	CYLKAKAATK
gi 1183031	SLTERLYVGG	PMYNSKGGSC	GYRRCRASGV	LPTSMGNLTL	CYLKAMAACK

	210	220	230	240	250
gi 329874:	AAGLQDCTML	VCGDDLVVIC	ESAGVQEDAA	SLRAFTEAMT	RYSAPPGDPP
gi 221611:	AAKLQDCTML	VNGDDLVVIC	ESAGTQEDAA	ALRAFTEAMT	RYSAPPGDPP
gi 221651:	AAGIAPTML	VCGDDLVVIS	ESQGTEEDER	NLRAFTEAMT	RYSAPPGDPP
gi 1183029	AANLRNPDFL	VCGDDLVVVA	ESDGVDEDRA	ALRAFTEAMT	RYSAPPGDAP
gi 2252490	AAALRDCTML	VCGDDLVVIA	ESDGVVEDNR	ALRAFTEAMT	RYSAPPGDAP
gi 2462304	AAKLRDCTLL	VCGDDLVAIC	ESQTHEDEA	SLRAFTEAMT	RYSAPPGDPP
gi 2326455	AANIKDCDML	VCGDDLVVIC	ESAGVQEDTA	SLRAFTDAMT	RYSAPPGDAP
gi 1183033	AAGIKDPSFL	VCGDDLVVIA	ESAGIDEDKS	ALRAFTEAMT	RYSAPPGDPP
gi 1183031	AAGLKNFDM	VCGDDLVVIS	ESLGVSEDAS	ALRAFTDAMT	RYSAPPGDPE

	260	270	280	290	
gi 329874:	QPEYDLELIT	SCSSNVSAH	DGAGKRVYYL	TRDPTTPLAR	Aawe
gi 221611:	QPEYDLELIT	SCSSNVSAH	DASGKRVYYL	TRDPTTPLAR	Aawe
gi 221651:	RPEYDLELIT	SCSSNVSAV	GPQGRRRYYL	TRDPTPIAR	Aawe
gi 1183029	QATYDLELIT	SCSSNVSVAR	DDKGRRRYYL	TRDATTPLAR	Aawe
gi 2252490	QPAYDLELIT	SCSSNVSAH	DVTGKRVYYL	TRDPETPLAR	AVWE
gi 2462304	VPAYDLELVT	SCSSNVSVAR	DASGNRIYYL	TRDPQVPLAK	Aawe
gi 2326455	QPTYDLELIT	SCSSNVSAH	EGNGKRVYYL	TRDCTTPLAR	Aawe
gi 1183033	QPTYDLELIT	SCSSNVSAH	DGAGKRVYYL	TRDPETPLAR	Aawe
gi 1183031	HPEYDLEHIT	SCSSNVSAH	DHTGQRYYYL	TRDPTNVLAR	Aawe