

Table des matières.

REMERCIEMENTS.....	2
TABLE DES MATIÈRES.....	3
TABLE DES ILLUSTRATIONS.....	6
ABRÉVIATIONS.....	8
1 INTRODUCTION.....	9
2 RAPPELS BIBLIOGRAPHIQUES.....	11
2.1 LE VIRUS DE L'HÉPATITE C.....	11
2.1.1 <i>Identification.....</i>	11
2.1.2 <i>La particule virale.....</i>	12
2.1.3 <i>Organisation génomique du virus de l'hépatite C.....</i>	12
2.1.3.1 La région 5' non-codante.....	14
2.1.3.2 La région 3' non-codante.....	17
2.1.3.3 La polyprotéine et sa maturation.....	17
2.1.3.3.1 La protéine de capside.....	18
2.1.3.3.2 Les glycoprotéines d'enveloppe E1 et E2.....	19
2.1.3.3.3 La protéine p7.....	21
2.1.3.3.4 La protéine NS2.....	21
2.1.3.3.5 La protéine NS3.....	21
2.1.3.3.5.1 Le domaine protéase à sérine.....	21
2.1.3.3.5.2 Le domaine hélicase.....	23
2.1.3.3.6 La protéine NS4A.....	25
2.1.3.3.7 La protéine NS4B.....	25
2.1.3.3.8 La protéine NS5A.....	26
2.1.3.3.9 La protéine NS5B.....	27
2.1.4 <i>Variabilité génétique et classification génotypique.....</i>	29
2.1.5 <i>La réplication du VHC.....</i>	33
2.2 L'HÉPATITE C.....	35
2.2.1 <i>Epidémiologie.....</i>	35
2.2.2 <i>Modes de transmission.....</i>	36
2.2.3 <i>Physiopathologie de l'infection.....</i>	37
2.2.4 <i>Diagnostic.....</i>	37
2.2.5 <i>Traitement.....</i>	38
2.2.6 <i>Vaccination.....</i>	39
2.3 LES BASES DE DONNÉES BIOLOGIQUES.....	42
2.3.1 <i>Bases de séquences.....</i>	44
2.3.1.1 Bases de séquences nucléiques.....	44
2.3.1.1.1 Les bases DDBJ, EMBL et Genbank.....	44
2.3.1.2 Bases de séquences protéiques.....	45
2.3.1.2.1 La base SWISS-PROT et les bases associées.....	45
2.3.1.2.2 PIR-PSD et les bases associées.....	48
2.3.1.2.3 La base GenPept.....	50
2.3.1.2.4 Autres bases.....	50
2.3.2 <i>Bases de signatures protéiques.....</i>	50
2.3.2.1 PROSITE.....	51
2.3.2.2 ProDom.....	51
2.3.2.3 Pfam.....	52
2.3.2.4 DOMO.....	52
2.3.2.5 PRINTS et PRINTS-S.....	52
2.3.2.6 Blocks et Blocks+.....	52
2.3.2.7 InterPro.....	52
2.3.3 <i>Bases de structures tridimensionnelles.....</i>	54
2.3.3.1 La Protein Data Bank.....	54
2.3.3.2 Bases de données exploitant la PDB.....	55

2.3.3.2.1	SCOP.....	55
2.3.3.2.2	CATH.....	56
2.3.3.2.3	Le dictionnaire de domaine DALI et la base FSSP.....	56
2.3.3.2.4	PALI.....	57
2.3.3.2.5	Bases de séquences.....	57
2.3.4	<i>Autres bases.....</i>	58
2.3.5	<i>Outils d'interrogation des bases de données.....</i>	58
2.4	MÉTHODES ET OUTILS DE LA BIOINFORMATIQUE.....	58
2.4.1	<i>Recherche d'homologie.....</i>	59
2.4.1.1	Par comparaison de séquences.....	59
2.4.1.1.1	Vocabulaire.....	59
2.4.1.1.2	Principe de l'alignement de deux séquences.....	59
2.4.1.1.2.1	Théorie.....	60
2.4.1.1.2.2	Matrice de substitution.....	61
2.4.1.1.2.3	Pénalisation des insertions et délétions.....	62
2.4.1.1.2.4	Signification statistique des alignements.....	62
2.4.1.1.3	Application à la recherche de similarité de séquences.....	63
2.4.1.1.3.1	Algorithmes.....	63
2.4.1.1.3.2	FASTA.....	64
2.4.1.1.3.3	BLAST.....	65
2.4.1.2	Par détection de motifs.....	66
2.4.1.2.1	Expressions régulières.....	66
2.4.1.2.2	Profils.....	66
2.4.1.3	Par l'utilisation de la structure tridimensionnelle des protéines.....	67
2.4.2	<i>Alignement multiple de séquences et phylogénie.....</i>	67
2.4.2.1	Algorithmes d'alignement multiple.....	67
2.4.2.1.1	L'algorithme CLUSTAL W.....	68
2.4.2.1.2	L'algorithme MultAlin.....	68
2.4.2.1.3	Problèmes liés aux alignements progressifs.....	68
2.4.2.2	Application à la phylogénie.....	69
2.4.3	<i>Prédictions de la structure secondaire des protéines.....</i>	71
2.4.3.1	Méthodes de statistiques linéaires.....	71
2.4.3.1.1	La méthode de Chou et Fasman.....	71
2.4.3.1.2	Les méthodes GOR.....	71
2.4.3.1.3	La méthode DPM.....	72
2.4.3.1.4	La méthode DSC.....	72
2.4.3.1.5	La méthode PREDATOR.....	72
2.4.3.2	Méthodes basées sur l'homologie.....	72
2.4.3.2.1	La méthode de Levin.....	72
2.4.3.2.2	La méthode SIMPA.....	73
2.4.3.2.3	Les méthodes auto-optimisées SOPM et SOPMA.....	73
2.4.3.3	Méthodes d'apprentissage.....	73
2.4.3.3.1	PHD.....	73
2.4.3.3.2	HNN.....	74
2.4.3.4	Améliorations de la qualité de prédiction.....	74
2.4.4	<i>Analyses spécialisées de la séquence des protéines.....</i>	74
2.4.4.1	Profils physico-chimiques.....	74
2.4.4.2	Détection de motifs coiled-coils.....	74
2.4.4.3	Détection de motifs hélice-coude-hélice de fixation à l'ADN.....	75
2.4.4.4	Prédiction des segments transmembranaires.....	75
2.4.5	<i>Prédiction de la structure tertiaire des protéines.....</i>	76
2.4.5.1	Modélisation moléculaire par homologie.....	76
2.4.5.2	Le threading.....	77
2.4.6	<i>Disponibilité des outils bioinformatiques.....</i>	77
2.5	LE WEB.....	78
3	MATÉRIELS ET MÉTHODES.....	80
3.1	MATÉRIELS INFORMATIQUES.....	80
3.2	LANGAGES DE PROGRAMMATION.....	81
3.2.1	<i>Le C.....</i>	81
3.2.2	<i>Le PERL.....</i>	82
3.2.3	<i>Icarus.....</i>	86
3.2.4	<i>HTML.....</i>	87
3.3	PROGRAMMES BIOINFORMATIQUES.....	87
3.4	PROGRAMMES INFORMATIQUES.....	88

4	RÉSULTATS ET DISCUSSION.....	90
4.1	LE SERVEUR NPS@	90
4.1.1	<i>Caractéristiques principales</i>	90
4.1.2	<i>Aspects techniques</i>	95
4.1.2.1	Le module pack_general.pm.....	95
4.1.2.2	Le module pack_simsearch.pm.....	96
4.1.2.3	Le module pack_pattern.pm.....	97
4.1.2.4	Le module pack_align.pm.....	99
4.1.2.5	Le module pack_secpred.pm.....	100
4.1.2.6	Le module pack_d3.pm.....	101
4.1.2.7	Le module pack_primanal.pm.....	102
4.1.2.8	Le module pack_db.pm.....	102
4.1.2.9	Les divers utilitaires CGI.....	104
4.1.3	<i>Statistiques d'utilisation</i>	104
4.1.4	<i>Bilan et perspectives</i>	105
4.2	LA BANQUE DE DONNÉES DE SÉQUENCES DU VHC : HCVDB.....	108
4.2.1	<i>Cahier des charges</i>	109
4.2.2	<i>Le module HCVWEB</i>	111
4.2.3	<i>Le module HCVDB</i>	111
4.2.3.1	Principe de la création de la banque.....	111
4.2.3.2	Annotation automatique des séquences.....	112
4.2.3.3	Génotypage.....	113
4.2.3.4	Les principaux champs de la banque HCVDB.....	114
4.2.3.4.1	Le champ SGACCN.....	115
4.2.3.4.2	Les champs SGCLAD et SGSUBT.....	115
4.2.3.4.3	Les champs SGXREF.....	115
4.2.3.4.4	Le champ SQCOMM.....	115
4.2.3.4.5	Le champ SQFEAT.....	115
4.2.4	<i>Le module HCVSRS</i>	115
4.2.5	<i>Le module HCVSA</i>	120
4.2.6	<i>Le module HCVFORM</i>	120
4.2.6.1	Le formulaire utilisateur.....	121
4.2.6.2	Le formulaire patient.....	121
4.2.6.3	Le formulaire clinique.....	121
4.2.7	<i>Bilan et perspectives</i>	122
5	CONCLUSION.....	124
6	RÉFÉRENCES.....	125
6.1	RÉFÉRENCES BIBLIOGRAPHIQUES.....	125
6.2	RÉFÉRENCES PERSONNELLES.....	156
6.2.1	<i>Publications</i>	156
6.2.2	<i>Communications</i>	156
6.2.3	<i>Séminaires</i>	157
6.2.4	<i>Posters</i>	157
6.3	ADRESSES WEB.....	159
6.3.1	<i>Bases de données</i>	159
6.3.2	<i>Outils bioinformatiques</i>	159
6.3.3	<i>Outils informatiques</i>	160
6.3.4	<i>Sites majeurs de BioInformatique</i> :	160
6.3.5	<i>Divers</i>	160
7	ANNEXES.....	161