

# **Bioinformatique avec Python et Biopython**

**Emmanuel Jaspard**  
Université d'Angers

**Gilles Hunault**  
Université d'Angers

**DUNOD**

© Dunod, Paris, 2024  
11, rue Paul Bert, 92240 Malakoff  
[www.dunod.com](http://www.dunod.com)  
ISBN 978-2-10-086862-9

À Gwilherm Jaspard.

Pour Raphaël et Timothée.



# Table des matières

|                                                                              |    |
|------------------------------------------------------------------------------|----|
| Introduction                                                                 | 1  |
| <b>Chapitre 1</b> Dénombrement des nucléotides et calculs associés (%GC, Tm) | 7  |
| 1. Rappel sur la structure des nucléotides                                   | 7  |
| 2. La composition en nucléotides de l'ADN et des ARN                         | 8  |
| 3. Illustration des îlots CpG                                                | 9  |
| 4. Calculs du pourcentage de GC et de la température de fusion Tm            | 10 |
| 5. Buts du script gctm.py                                                    | 11 |
| 6. Code du script gctm.py                                                    | 12 |
| 7. Explications du code gctm.py                                              | 12 |
| 8. Code de l'auto-test pour le script gctm.py                                | 14 |
| 9. Calculs de %GC et de Tm avec Biopython                                    | 15 |
| L'essentiel                                                                  | 16 |
| Bibliographie                                                                | 16 |
| Ressources Web                                                               | 16 |
| Exercices                                                                    | 18 |
| <b>Chapitre 2</b> Écriture de séquences d'acides nucléiques                  | 19 |
| 1. Les séquences d'acides nucléiques                                         | 19 |
| 2. Les formats des fichiers bioinformatiques – le format FASTA               | 20 |
| 3. Buts du script brinscomp.py                                               | 22 |
| 4. Code du script brinscomp.py                                               | 23 |
| 5. Explications du code brinscomp.py                                         | 23 |
| 6. Code de l'auto-test pour le script brinscomp.py                           | 24 |
| L'essentiel                                                                  | 25 |
| Bibliographie                                                                | 25 |

|                                                                                          |    |
|------------------------------------------------------------------------------------------|----|
| Ressources Web                                                                           | 25 |
| Exercices                                                                                | 26 |
| <b>Chapitre 3 Séquences et expressions régulières</b>                                    | 27 |
| 1. Rappels sur les caractéristiques des expressions régulières                           | 27 |
| 2. Utilisation des expressions régulières dans l'analyse des données bioinformatiques    | 28 |
| 3. Buts du script <code>rchmotif.py</code>                                               | 29 |
| 4. Code du script <code>rchmotif.py</code>                                               | 30 |
| 5. Explications du code <code>rchmotif.py</code>                                         | 31 |
| 6. Code de l'auto-test pour le script <code>rchmotif.py</code>                           | 31 |
| L'essentiel                                                                              | 32 |
| Bibliographie                                                                            | 32 |
| Ressources Web                                                                           | 32 |
| Exercices                                                                                | 33 |
| <b>Chapitre 4 Recherches bibliographiques dans PubMed</b>                                | 35 |
| 1. Rappel sur les données bibliographiques et le moteur de recherche <code>Entrez</code> | 35 |
| 2. Buts du script <code>rchbiblio.py</code>                                              | 37 |
| 3. Code du script <code>rchbiblio.py</code>                                              | 38 |
| 4. Explications du code <code>rchbiblio.py</code>                                        | 39 |
| 5. Code de l'auto-test pour le script <code>rchbiblio.py</code>                          | 40 |
| L'essentiel                                                                              | 41 |
| Bibliographie                                                                            | 41 |
| Ressources Web                                                                           | 41 |
| Exercices                                                                                | 42 |

|                   |                                                                                      |    |
|-------------------|--------------------------------------------------------------------------------------|----|
| <b>Chapitre 5</b> | <b>Profil d'une famille de protéines</b>                                             | 43 |
| 1.                | Rappel sur les domaines structuraux des protéines                                    | 43 |
| 2.                | Profils, domaines et bases de données de familles de protéines                       | 44 |
| 3.                | Méthode de construction des alignements <b>PFAM</b>                                  | 45 |
| 4.                | Buts du script <b>familprot.py</b>                                                   | 45 |
| 5.                | Code du script <b>familprot.py</b>                                                   | 47 |
| 6.                | Explications du code <b>familprot.py</b>                                             | 48 |
| 7.                | Code de l'auto-test pour le script <b>familprot.py</b>                               | 51 |
|                   | L'essentiel                                                                          | 52 |
|                   | Bibliographie                                                                        | 52 |
|                   | Ressources Web                                                                       | 52 |
|                   | Exercices                                                                            | 53 |
| <b>Chapitre 6</b> | <b>Profil d'hydrophobicité d'une protéine –<br/>fenêtre de taille variable</b>       | 55 |
| 1.                | Propriétés physico-chimiques des acides aminés                                       | 55 |
| 2.                | Calcul de l'hydrophobicité d'une séquence d'acides aminés avec une fenêtre glissante | 56 |
| 3.                | Exemple d'un récepteur couplé à une protéine <b>G (RCPG)</b>                         | 57 |
| 4.                | Buts du script <b>profilhyd.py</b>                                                   | 57 |
| 5.                | Code du script <b>profilhyd.py</b>                                                   | 58 |
| 6.                | Explications du code <b>profilhyd.py</b>                                             | 59 |
| 7.                | Code de l'auto-test pour le script <b>profilhyd.py</b>                               | 60 |
| 8.                | Code du script <b>profilhyd_utils.py</b>                                             | 61 |
| 9.                | Explications du code <b>profilhyd_utils.py</b>                                       | 62 |
| 10.               | Calculs avec les fonctions de <b>Biopython</b>                                       | 62 |
|                   | L'essentiel                                                                          | 63 |
|                   | Bibliographie                                                                        | 63 |

|                                                                                                     |    |
|-----------------------------------------------------------------------------------------------------|----|
| Ressources Web                                                                                      | 63 |
| Exercices                                                                                           | 64 |
| <b>Chapitre 7 Position de cystéines impliquées dans un pont disulfure</b>                           | 65 |
| 1. Propriétés physico-chimiques et particularité de la cystéine                                     | 65 |
| 2. Pont disulfure et structure 3D des protéines                                                     | 66 |
| 3. La base de données « <i>Protein DataBank</i> » ou PDB                                            | 67 |
| 4. Buts du script <code>disulfure.py</code>                                                         | 69 |
| 5. Code du script <code>disulfure.py</code>                                                         | 70 |
| 6. Explications du code <code>disulfure.py</code>                                                   | 71 |
| 7. Code de l'auto-test pour le script <code>disulfure.py</code>                                     | 73 |
| L'essentiel                                                                                         | 74 |
| Bibliographie                                                                                       | 74 |
| Ressources Web                                                                                      | 74 |
| Exercices                                                                                           | 75 |
| <b>Chapitre 8 Calcul du barycentre d'une chaîne polypeptidique repliée</b>                          | 77 |
| 1. Structure tridimensionnelle des protéines                                                        | 77 |
| 2. Calcul du barycentre des carbones $\alpha$ des acides aminés d'une chaîne polypeptidique repliée | 78 |
| 3. Buts du script <code>barycentre.py</code>                                                        | 78 |
| 4. Code du script <code>barycentre.py</code>                                                        | 79 |
| 5. Explications du code <code>barycentre.py</code>                                                  | 80 |
| 6. Code de l'auto-test pour le script <code>barycentre.py</code>                                    | 82 |
| L'essentiel                                                                                         | 83 |
| Bibliographie                                                                                       | 83 |
| Ressources Web                                                                                      | 83 |
| Exercices                                                                                           | 84 |



|                    |                                                                      |     |
|--------------------|----------------------------------------------------------------------|-----|
| <b>Chapitre 9</b>  | <b>Diagrammes de Ramachandran</b>                                    | 85  |
| 1.                 | La liaison peptidique                                                | 85  |
| 2.                 | Les angles de torsion                                                | 86  |
| 3.                 | Les diagrammes de Ramachandran                                       | 87  |
| 4.                 | Buts du script <code>ramachandran.py</code>                          | 87  |
| 5.                 | Code du script <code>ramachandran.py</code>                          | 89  |
| 6.                 | Explications du code <code>ramachandran.py</code>                    | 90  |
| 7.                 | Code de l'auto-test pour le script <code>ramachandran.py</code>      | 91  |
| 8.                 | Code du script <code>ramachandran_utils.py</code>                    | 92  |
| 9.                 | Explications du code <code>ramachandran_utils.py</code>              | 94  |
|                    | L'essentiel                                                          | 96  |
|                    | Bibliographie                                                        | 96  |
|                    | Ressources Web                                                       | 96  |
|                    | Exercices                                                            | 97  |
| <b>Chapitre 10</b> | <b>Contacts entre résidus d'acides aminés dans les structures 3D</b> | 99  |
| 1.                 | Interactions entre résidus d'acides aminés au sein des protéines     | 99  |
| 2.                 | Contacts et distances entre résidus d'acides aminés                  | 100 |
| 3.                 | Buts du script <code>contacts.py</code>                              | 101 |
| 4.                 | Code du script <code>contacts.py</code>                              | 102 |
| 5.                 | Explications du code <code>contacts.py</code>                        | 102 |
| 6.                 | Code de l'auto-test pour le script <code>contacts.py</code>          | 103 |
| 7.                 | Code du script <code>structure_pdb.py</code>                         | 103 |
| 8.                 | Explications du code <code>structure_pdb.py</code>                   | 104 |
| 9.                 | Code du script <code>contacts_utils.py</code>                        | 105 |
| 10.                | Explications du code <code>contacts_utils.py</code>                  | 106 |

|                                                                                   |     |
|-----------------------------------------------------------------------------------|-----|
| L'essentiel                                                                       | 108 |
| Bibliographie                                                                     | 108 |
| Ressources Web                                                                    | 108 |
| Exercices                                                                         | 109 |
| <b>Chapitre 11 Superposition de structures protéiques</b>                         | 111 |
| 1. Superposition de structures de protéines et scores de distances interatomiques | 111 |
| 2. Exemples de scores de mesure de similarité de structures                       | 112 |
| 3. L'algorithme <b>AlphaFold</b>                                                  | 113 |
| 4. Buts du script <b>superp.py</b>                                                | 114 |
| 5. Code du script <b>superp.py</b>                                                | 115 |
| 6. Explications du code <b>superp.py</b>                                          | 116 |
| 7. Code de l'auto-test pour le script <b>superp.py</b>                            | 116 |
| 8. Code du script <b>superp_pdb.py</b>                                            | 117 |
| 9. Explications du code <b>superp_pdb.py</b>                                      | 118 |
| L'essentiel                                                                       | 120 |
| Bibliographie                                                                     | 120 |
| Ressources Web                                                                    | 120 |
| Exercices                                                                         | 121 |
| <b>Chapitre 12 Les protéines ou régions intrinsèquement désordonnées</b>          | 123 |
| 1. Découverte des protéines ou régions intrinsèquement désordonnées               | 123 |
| 2. Abondance et fonctions biologiques des <b>IDP/IDR</b>                          | 125 |
| 3. Prédiction des <b>IDP/IDR</b>                                                  | 126 |
| 4. Buts du script <b>repliement.py</b>                                            | 126 |
| 5. Code du script <b>repliement.py</b>                                            | 127 |