

Introduction

Cette étude se consacre sur l'analyse de la reconnaissance de modèles des bonds d'hydrogènes et des caractéristiques géométriques des protéines. Autrement dit, nous nous intéressons à l'assignation des structures secondaires de protéines, à comprendre de quels motifs ces dernières sont composées. Ce rapport se base sur un article scientifique nommé *Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features*, Kabsch W, Sander C, 1983. Cinq protéines issues de cet article ont été choisies. Grâce à la méthode DSSP (Hydrogen Bond Estimation Algorithm), la structure de protéines sélectionnées pourra alors être déterminée et les résultats obtenus seront comparés avec ceux présents dans l'article cité précédemment. Le travail a été géré à distance via [Trello](#) pour l'organisation, [github](#) et zoom.

Matériels et Méthodes

Les fichiers pdb des protéines choisies ont été téléchargés ([Protein Data Bank](#)): [1EST](#) (hydroxylase), [2SGA](#) (anciennement 1SGA, protéase), [2MHB](#) (hémoglobine), [6LDH](#) (4LDH, oxydoréductase), [7FAB](#) (1FAB, immunoglobuline). Le choix de ces protéines est basé sur le but d'avoir des protéines avec plusieurs types de structures secondaire. Protéine constitué que d'hélice, que de feuillet beta ou des protéines a structures mixtes. La longueur de ces liaisons peptidiques se situent entre 2.2 et 3 Angstrom. Il a fallu dans un premier temps vérifier ces distances à l'aide d'un 1er script ([verification_liaison_peptidique.py](#)). Par la suite, notre but est de connaître les structures secondaires de ces protéines et donc ainsi connaître la position des différents atomes les composant. Cependant, les atomes d'hydrogènes (H) sont absents dans les fichiers pdb. Pour ce faire, le package [Reduce](#) a été utilisé afin de les rajouter en utilisant un script python [pdbHydrogene.py](#). Le fichier pdb issu de ce script contient donc les atomes H. Il servira donc d'input pour le script [parser_fichier.py](#) qui a pour objectif de déterminer les atomes d'intérêt à une liaison Hydrogène. Dans la logique de notre conception, le fichier issue à cet étape, est sensé être l'input à la prochaine étape. Mais à cause de problèmes rencontrés, le fichier a subi quelques modifications avant d'être utilisé par le script [calc_energie.py](#) de calcul de distances entre les atomes et le calcul de l'énergie de liaison qui définit le type de structures secondaires. Une comparaison des résultats obtenue dans ce travail et les résultats de l'article de départ a été réalisé. La partie étudié ici est celle des Hélices, l'assignation des feuillets beta n'a pas été établie.

Résultats et Observations

Suite aux différentes étapes effectuées dans la partie Matériels et Méthodes, des fichiers output sous format texte (.txt) ont pu être obtenus pour nos 5 protéines d'intérêt. Ils récapitulent le couple de résidu potentiellement impliqués dans la liaison d'hydrogène, les distances entre les 4 atomes de la liaison peptidique (N,C,O,H), l'énergie de liaison calculée et si cette énergie est inférieure à -0.5 kcal/mole, la lettre 'H' sera notée (Annexe 1). La présence de cette lettre signifiera la présence d'une liaison hydrogène entre les deux résidus étudiés ($Hbond(i,j) = [E < -0.5 \text{ kcal/mole}]$ avec i et j deux résidus distincts).

De ces liaisons hydrogènes, les hélices alpha peuvent être déduites si deux lettres H se suivent dans nos fichiers. Ces hélices sont possibles grâce à des liaisons hydrogènes entre le résidu en position i et le résidu en position i+4. Le nombre d'hélices alpha présentes dans chacune des protéines étudiées a pu être déterminé (NB: la même démarche a pu être effectuée pour calculer le nombre des hélices 3.10 ($i+3$) et π ($i+5$), $n\text{-turn}(i) = Hbond(i, i+n)$, $n = 3,4,5$). Ces valeurs ont pu être notées dans le Tableau 1. On peut voir que nous obtenons le même ordre de grandeur d'hélices entre nos résultats et les données issues de la base de donnée Protein Data Bank.

La recherche des feuillets bêta n'a pas pu être réalisée suite à une gestion de temps.

Protéine choisie	Hélices 3.10	Structure secon	daires (hélices)		Feuillet(PDB)
		Hélices alpha	Hélices pi	Helices (PDB)	
1EST	0	0	0	2	14
2MHB	1	9	0	8	0
2SGA	0	0	0	2	21
6LDH	1	8	0	11	13
7FAB	0	0	0	2	9

Tableau 1 - Récapitulatif du nombre de chacune des hélices dans les protéines suite à notre analyse et issu de la base de données PDB.

Conclusion

L'assignation des structures secondaires en s'appuyant sur les liaisons hydrogènes est une procédure assez minutieuse car elle se fait sur des calculs de distances et d'énergie. La principale contrainte est la qualité des fichiers pdb, il n'est pas rare de trouver des fichiers pdb contenant des erreurs ou des imprécisions. Le passage par reduce est primordial et le résultat peut donc varier selon la version du reduce. Ensuite, nous n'avons pas réussi à détecter les feuillets bêta mais cela sera une suite logique à cet étude. Une fois ces deux grandes classes de structures secondaires détectées on pourrait passer à une assignation plus détaillée voir les types d'hélices etc. Les résultats obtenus dans ces études ne sont pas assez exploitables et comparables à ceux de l'étude de départ.

ANNEXES

Annexe I - Photo montrant les résultats obtenus à la fin de notre programme pour deux protéines (1EST (A)et 2MHB)

A. 1EST

 test_H_parse_helices_alpha - Bloc-notes

Fichier Edition Format Affichage Aide

n° residus	Distance_NO	Distance_HC	Distance_HO	Distance_CN	Energie E>-0.5kcal/mol
(17, 21)	12.630656712934604	12.084417900751365	12.517809512850087	11.964382725406269	-0.04305775827986581
(21, 25)	13.263091230931042	9.135193758207865	12.018163670045436	10.396350994459542	0.15251877547805634
(25, 30)	11.645714791286965	11.927903294376595	12.949607291342856	10.61398987186251	-0.0483067690549102
(30, 34)	15.997958244726112	14.278161506300451	17.077289656148604	13.161285423544314	-0.05557289632285781
(34, 36)	13.606707463600442	11.878819006955197	13.774516543240273	11.51759953288879	-0.04866061320545123
(36, 39)	9.269289616793726	12.251903239905218	10.51084249715502	11.084961028348273	0.11576133560511292
(39, 43)	13.074365185354127	11.558459456173214	13.929639837411449	10.602511494924212	-0.08657472023541342
(43, 47)	15.837997663846274	13.231493113023939	15.877506510784366	13.243378911743028	0.006273202133424438
(47, 51)	3.302238331798597	7.066269595762674	3.8868214520350657	6.035270582169449	0.5959665649681859
(51, 55)	15.413697058136313	13.739826672851446	15.868238906696611	13.195440197280272	-0.03191046188000625
(55, 59)	8.67337829222704	6.084925143335784	8.872068191802859	5.913477065145345	-0.060869742170346006
(59, 63)	9.927009720958269	9.13997007653745	10.86623587080641	7.940813875164182	-0.21794600561089653
(63, 66)	14.411289116522504	13.7046994129751	14.931044638604495	13.023219494426101	-0.039120242974679455
(66, 70)	12.772878845428702	11.35260749783943	13.561801207804221	10.540152418252784	-0.06234167335520987
(70, 74)	11.721893244566822	9.958194063182336	10.726193033877397	11.0227075167583	0.04960525001357
(74, 78)	12.6309347635082	12.13905226943191	11.953480539156786	12.41906059249249	-0.07333351779206772
(78, 82)	12.469668840831337	13.21709389389362	13.332789243065383	12.322671017275434	-0.008369444259382009
(82, 86)	15.782373902553443	13.418913331563031	15.850954356126321	13.220388004896076	-0.023563197096956322
(86, 90)	14.538636077796876	14.438548853676396	15.308992945324654	13.402581579680833	-0.05277228093342506
(90, 94)	8.72673902440081	10.922346405420402	11.836270738708203	7.364759466540642	-0.3938362897075907
(94, 98)	5.052531444731443	10.21149514028186	7.608703700368414	8.199289542393291	1.1841028007868215
(98, 100)	5.625270571270328	9.249236238738852	8.317754865346776	6.83460445088083	0.5395565392263937
(100, 104)	11.288639156249081	17.57135996444214	14.38482415603333	14.407675211497516	0.18323169802305228
(104, 108)	12.106883909578055	18.04821082545303	14.901973090835991	15.40264438335184	0.16664939041648782
(108, 112)	10.881254569212137	13.325287013794485	13.76885634321166	10.41567981458724	-0.0471394624766212
(112, 116)	11.68966013192856	11.19622257728025	9.060659633823576	14.093204710072161	-0.18020941892271303
(116, 120)	10.395061760278292	14.94485761062904	13.593957959328842	11.545567591071478	0.08190021431880662
(120, 125)	14.631709161953705	20.33837604136574	17.09466975404907	17.86749218553068	0.08498923079020734
(125, 129)	10.373747297866862	11.598420323475091	13.209128245270392	9.17404643546129	-0.05835726642612561
(129, 133)	10.183641097367872	12.345845536049767	8.625707217382235	13.914896011109818	-0.23990210337806647
(133, 138)	14.824402753568187	18.78312995749111	17.86667165422816	15.807925164296549	0.04088559337835145
(138, 142)	10.101003960003181	12.764359835103363	11.59851417208256	12.078301370639826	0.23236724675657733
(142, 147)	8.900817996116986	12.146738533450039	7.988821252224886	12.90583623791965	-0.22264067052145664
(147, 151)	11.158644854999196	13.481827843434287	13.584172260391872	11.649849312330181	0.12096224373299079
(151, 155)	8.763851322335404	12.832694027366191	8.941510722467427	12.975040423829132	0.08706827903873168
(155, 159)	12.939335608909753	18.535576333095225	14.783601252739468	16.802028091870334	0.11364037032148402
(159, 164)	14.84054534038423	21.309345391165824	17.53911699601779	18.572072636084542	0.09624249802707362
(164, 168)	7.255961204416679	10.142932514810497	7.947134829609976	8.776649531569548	-0.09375042406340216
(168, 171)	8.553722347609837	7.627515847246729	6.046898130446717	9.537956437308775	-0.6192769014503526 H
(171, 174)	11.38633558261832	11.961393104484111	10.6521075379476	11.93368249954724	-0.17423577437358947
(174, 178)	11.59787019241033	13.649236681954052	11.415517596675151	13.272494076095873	-0.09640740924945575
(178, 182)	11.401809330101957	15.802121692987939	14.833214418999008	12.470484593631475	0.09432845864457816
(182, 186)	12.349269573541589	16.212884567528384	12.61532151790037	15.678498142360446	-0.011002387892917768
(186, 189)	9.249192451235945	10.790184660143684	10.932520203502941	8.972811265149849	-0.05922413150488877
(189, 193)	10.026244561150502	11.250576029697326	8.128115033634797	12.620008914418406	-0.38057142887816703
(193, 197)	8.03567769388494	15.344775299755936	13.112985548684168	9.923344244759425	0.3508591960971083
(197, 202)	12.385266287004088	21.023174474850364	14.162163782416867	19.415989003911182	0.1727112592307615
(202, 206)	2.8367216994269997	6.385878404730236	5.730648305383955	5.215787093047413	3.984893580832518
(206, 210)	12.526929751539283	16.029792543885275	13.450252413988373	14.266817094222523	-0.0621597505352271
(210, 214)	12.380435250830235	15.797853588383452	13.136233782937936	15.20317391204876	0.060552765529634116
(214, 217)	12.116882808709507	15.453890222206185	12.986859089094636	14.891134443016758	0.08598259138118466
(217, 221)	8.945808292155606	9.91055528212219	9.034176940928267	7.5909931451010341	0.21126682226650662
(221, 224)	4.6736825951277465	10.6816595152626	9.799974438742176	7.0006024026507925	1.7484804752786134
<					
					Ln 2, Col 1

B. 2MHB

 2mhb_H_parse_helices_alpha - Bloc-notes

Fichier Edition Format Affichage Aide

n° residus	Distance_NO	Distance_HC	Distance_HO	Distance_CN	Energie E>-0.5kcal/mol
(3, 7)	3.0109270997485136	3.3817168716496657	2.207551584901245	4.212003205126986	-1.7451158548603856 H
(7, 11)	2.98544653276524	3.2675825926822415	2.1475953995108106	4.117720850179137	-1.8822988386244437 H
(11, 15)	3.2297916341460793	3.6002434639896226	2.552709540860456	4.346498705855093	-0.9603055157841179 H
(15, 19)	5.927691877957219	6.641704675759079	5.732424094569416	6.7326655939531115	-0.10353053014361586
(19, 23)	5.239308733029578	5.042446727532181	4.3994137109392195	5.872461494126633	-0.23448403354002878
(23, 27)	3.0491774300620835	3.32308681800521	2.2202497607251312	4.172309911787474	-1.7065458166217742 H
(27, 31)	2.966206499891739	3.3441916811092036	2.231447288196608	4.128487132110261	-1.5115925938383856 H
(31, 35)	2.9041537493734726	3.12406369973469	2.0865322906679395	3.975933852568475	-1.8502861502138819 H
(35, 40)	9.522041062713393	8.30775426935583	8.933321666659047	8.83342974161226	0.006754953257206217
(40, 45)	6.80351276915095	7.829077915054875	6.617273381083784	8.037179107124588	-0.02313451735003396
(45, 49)	9.687932596792775	9.007905194882992	9.04451994303733	9.595007087021877	-0.015345656657476853
(49, 53)	6.493294849304166	7.254163080604128	6.609896897834338	7.251298159640108	0.0742452553588629
(53, 57)	2.981500628878015	3.1530031715810254	2.248444128725462	3.955436006308281	-1.2552138843416571 H
(57, 61)	3.394456804851111	3.6448334667032443	2.672925363716691	4.429978103783358	-0.8616772555102004 H
(61, 65)	3.0083964166977712	3.2686111729601626	2.2459321004874564	4.080938740044992	-1.4487190456582755 H
(65, 69)	3.11313395150289	3.3426752758830762	2.3493797053690573	4.156578641142255	-1.278538096706653 H
(69, 73)	2.767181237288227	3.5251236857733095	2.4114794214340693	3.9546309562334656	-0.6273331561849554 H
(73, 77)	6.800625412416125	7.908124556429293	7.902311180914101	7.134425414845962	0.1892707527039982
(77, 82)	5.692865447206707	5.94801571954883	5.053152085579851	6.535541140563651	-0.19867520848718145
(82, 86)	3.203573785633788	3.5457057125486324	2.4080948901569488	4.376037134211729	-1.3832613986000353 H
(86, 90)	3.2564492933254767	3.3869046340279496	2.65828422107193	4.129887891940894	-0.44570726358094953
(90, 94)	9.325400473974293	9.219026683983511	9.166774841785958	9.294404445686663	-0.027216375277603958
(94, 99)	6.469225687823852	6.603908312507071	5.798525847144255	7.3307434820760164	-0.07992562606590035
(99, 103)	3.0361114933414415	3.3181307991096434	2.2593273335220823	4.13079374939006	-1.5045763762381685 H
(103, 107)	2.918822536571897	3.210904856890033	2.1577546199695656	4.032371758655197	-1.6006356528158148 H
(107, 111)	3.172037042658865	3.4795531322283306	2.4568559176313114	4.258747233635731	-1.092848602323122 H
(111, 116)	8.974195785695787	8.25967965480502	8.196117983533423	9.036470439281036	-0.004768517577721567
(116, 121)	5.668098093717151	6.138923765612342	5.189019464214795	6.6830763874132115	-0.08436956832925788
(121, 125)	3.023015216633882	3.342809895881007	2.293647313777773	4.132331182274721	-1.3396250112723957 H
(125, 129)	2.9639986504720284	3.2986239555305468	2.200569698964338	4.098418597459269	-1.61431207423784 H
(129, 133)	3.0698555340601947	3.2762168731633134	2.2954831735388526	4.102065455352949	-1.350878957685702 H
(133, 137)	2.999781992078758	3.5375331800564083	2.4604477234844886	4.163915224881507	-0.8519299536289383 H
(137, 141)	3.6292040449663343	3.9028699696505393	2.8440444792583683	4.699185567734052	-0.9105587065857459 H
(141, 5)	43.82950089836753	43.65447169534869	43.894931313307694	43.56589456214574	-0.0003504100904366978
(5, 9)	3.4405508279925194	3.4052110066778525	2.6320687680985855	4.258766957700315	-0.8483624452556908 H
(9, 13)	3.202828281378819	3.3953478466867	2.4203460496383578	4.223069144591408	-1.205151718671481 H
(13, 17)	4.264710541173925	4.406444031188866	3.7642088146116435	5.023496093359682	-0.09207960955589681
(17, 21)	9.007473841205424	8.83005345397184	8.573979414484269	9.269088088911444	-0.006941872876216716
(21, 25)	3.0454369144672815	3.3153233628109335	2.2569195820852794	4.140311220186232	-1.5232349918320902 H
(25, 29)	2.9870349847298425	3.259173054625973	2.2740178099566415	4.057132731375695	-1.2444557574843391 H
(29, 33)	3.2338803317377094	3.726677072138127	2.616923002306334	4.412873213678363	-0.8694445490799405 H
(33, 38)	8.759803365372994	7.581053884520278	7.985363235319981	8.357004786405236	0.03280688476223099
(38, 42)	4.108768306926055	4.507542345890939	3.5958527500441386	5.1349467377958256	-0.21222249314377606
(42, 46)	4.381187167880414	4.805416631261019	3.9083432551402115	5.388388720201989	-0.14222842567610097
(46, 50)	8.424732043216569	8.159287346331173	7.840497433198993	8.82399036717516	0.010807146407738649
(50, 55)	5.097668290503021	5.349717282249597	4.435821569901115	6.0614966798638115	-0.20411933299851157
(55, 60)	7.130478314951952	6.902498895327692	6.699777384361364	7.408822376059506	0.024686650808143223

<

Ln 1, Col 82 90% Unix (I