



Institut de XXX

École Doctorale de XXX

# Titre de la thèse

## THÈSE

présentée et soutenue publiquement le XXX

pour l'obtention du

**Doctorat de l'Université XXX**

(spécialité XXX)

par

David MIGNON

### Composition du jury

Rapporteurs : Dr. Prénom1 NOM1 Rapporteur externe  
Dr. Prénom2 NOM2 Rapporteur externe  
Pr. Prénom3 NOM3 Rapporteur interne

Examinateurs : Dr. Prénom4 NOM4 Examinateur  
Dr. Prénom5 NOM5 Directeur de thèse  
Dr. Prénom6 NOM6 Directeur de thèse

---

Laboratoire de XXX



# Remerciements

XXX



*à XXX.*



# Table des matières

Liste des figures	vii
Liste des tables	ix
Abreviations	xi
<b>Introduction</b>	<b>1</b>
0.1    Section . . . . .	3
0.1.1    Subsection . . . . .	3
<b>1    Methodes : La theorie</b>	<b>5</b>
1.1    Section . . . . .	5
<b>2    Les Méthodes utilisées</b>	<b>7</b>
2.1    Section . . . . .	7
2.1.1    Subsection . . . . .	7
<b>3    Les comparaisons d'algorithmes</b>	<b>9</b>
3.1    Les méthodes pratiques . . . . .	9
3.1.1    Protocoles Monte-Carlo . . . . .	13
3.1.2    Les protocoles Replica Exchange . . . . .	14
3.1.3    Les protocoles Toulbar2 . . . . .	16
3.1.4    Les temps de calculs . . . . .	17
3.2    Les tests . . . . .	18
3.2.1    Tous les résidus actifs . . . . .	18
3.2.2    Tous les résidus inactifs . . . . .	22
3.2.3    Etude au voisinage de GMECs . . . . .	23
3.2.4    Résultats Superfamily . . . . .	37
3.2.5    Résultats Heuristic (protocoles longs) . . . . .	38

## **Table des matières**

---

3.2.6 densité en séquences . . . . .	39
<b>Conclusion</b>	<b>41</b>
<b>Bibliographie</b>	<b>45</b>

# Liste des figures

3.1	Temps d'occupation du processeur selon le nombre de positions actives. . . . .	17
3.2	Distribution des énergies selon la température (protocole RE3). . . . .	18
3.3	Variation de la température au court de la trajectoire de chaque marcheur (protocole RE1). . . . .	19
3.4	Tous les protocoles. . . . .	20
3.5	Variation de la température au court de la trajectoire de chaque marcheur (protocole RE1). . . . .	21
3.6	test : 1M61 2, GMEC vs H . . . . .	34
3.7	test : 1M61 2, GMEC vs RE . . . . .	34
3.8	test : 1G9O 3, GMEC vs H . . . . .	35
3.9	test : 1G9O 5, GMEC vs MC . . . . .	35
3.10	test : 1CKA 5, GMEC vs H . . . . .	36
3.11	test : 1CKA 5, GMEC vs MC . . . . .	36
3.12	Alignement du voisinage : 1CKA 5 . . . . .	37
3.13	. . . . .	39
3.14	. . . . .	39



# Liste des tables

3.1	Les protéines . . . . .	10
3.2	Les protocoles Heuristique . . . . .	13
3.3	Les protocoles Monte-Carlo . . . . .	15
3.4	Les protocoles Replica Exchange . . . . .	16
3.5	les meilleures énergies pour tous les résidus actifs . . . . .	18
3.6	L'énergie du GMEC et la différence avec les autres protocoles. Tous les résidus inactifs () . . . . .	22
3.7	Liste des échecs pour 1A81 . . . . .	22
3.8	Liste des échecs pour 1BM2 . . . . .	24
3.9	Liste des échecs pour 1CKA . . . . .	25
3.10	Liste des échecs pour 1M61 . . . . .	25
3.11	Liste des échecs pour 1O4C . . . . .	26
3.12	Liste des échecs pour 1R6J . . . . .	27
3.13	Liste des échecs pour 2BYG . . . . .	27
3.14	Résultats 5 position actives . . . . .	28
3.15	Résultats 10 positions actives . . . . .	29
3.16	Mutations 10 positions actives . . . . .	30
3.17	Résultats pour 3 fois (resp 9 fois)plus de cycles heuristiques protocole H+ (resp H++) . . . . .	38



# Abbreviations

**H** algorithme heuristique

**MC** algorithme Monte-Carlo

**RE** algorithme “Replica Exchange”

**GMEC** “Global minimal energie cost”



# Introduction

XXX



# Contexte

## 0.1 Section

XXX

### 0.1.1 Subsection

XXX

#### Subsubsection

XXX

##### Paragraph XXX

###### Subparagraph XX

Citation entre crochets [Kofke 2002, 2004a].

Citation dans le texte Kofke [2004b].



## Chapitre 1

# Methodes : La theorie

## 1.1 Section

Subparagraph



## Chapitre 2

# Les Méthodes utilisées

## 2.1 Section

XXX

### 2.1.1 Subsection

XXX

#### Subsubsection

XXX

**Paragraph** XXX

**Subparagraph** Référence à une équation (Équation 2.1).

$$H\Psi = E\Psi \tag{2.1}$$



## Chapitre 3

# Les comparaisons d'algorithmes

### 3.1 Les méthodes pratiques

Nous cherchons maintenant à déterminer les performances et les qualités des différents algorithmes de proteus. Pour évaluer les qualités des différents algorithmes de proteus, nous effectuons un ensemble de tests. Plusieurs questions se présentent alors, premièrement grâce à l' algorithme de type toulbar2 il est possible d'obtenir la séquence/conformation qui possède l'énergie de déploiement la plus haute. Cela constitue une information importante qui va nous servir de point de comparaison. Mais le facteur temps pose problème, nous ne savons pas à l'avance quand toulbar2 termine. Et il apparaît d'emblée illusoire d'espérer voir toulbar2 converger dans toutes les situations qui nous intéressent dans un temps raisonnable, en particulier pour une situation où toutes les positions du backbone sont autorisé à muter et pour chaque type d'acide aminé la chaîne latérale peut prendre toutes les positions précédemment calculées dans la matrice.

Dans la suite on appelle position active, une position pour laquelle, lors de la recherche par proteus, tous les types d'acides aminés sont autorisés et tous les rotamères de chaque type d'acide aminé sont autorisés.

#### les protéines

Les tests sont effectués sur neuf protéines choisies pour avoir des longueurs de backbone variés , plusieurs familles SCOP(?) représentées, mais aussi plusieurs structures pour chaque famille présente. Ainsi l'ensemble se décompose en deux protéines SH3 de 56 et 57 résidus, de trois protéines PDZ de longueur comprise entre 82 et 97 résidus et enfin de trois protéines SH2 longues de 105 ou 109 résidus (Table 3.1).

#### Description des tests

Les tests vont donc être organisés en deux ensembles :

### **Chapitre 3. Les comparaisons d’algorithmes**

---

Code PDB	résidus	nombre de positions	famille
1ABO	64-119	56	SH3
1CKA	134-190	57	SH3
1R6J	192-273	82	PDZ
1G9O	9-99	91	PDZ
2BYG	186-282	97	PDZ
1BM2	55-152	98	SH2
1O4C	1-105	105	SH2
1M61	4-112	109	SH2
1A81	9-117	109	SH2

Table 3.1 – Les protéines

1. un ensemble de tests où toutes les positions du backbone sont actives (cela correspond aux situations de design complet de protéines)
2. un ensemble de tests où le nombre de positions actives est gardé sous contrôle .De façon à maîtriser la taille de l'espace d'exploration

**L’ensemble ’Tout actif’** Pour le premier ensemble de tests,La totalité de la matrice d’énergie est exploitée et pour chaque position l’espace d’exploration correspond à l’espace d’état déclaré dans le fichier .bb. Comme l’espace des séquences/confirmations à explorer est gigantesque, nous ne faisons pas de tentative de recherche du GMEC par recherche exacte.

Nous effectuons des recherches avec les algorithmes suivants :

- heuristique (noté H par la suite) ;
- Monte-Carlo (noté MC) ;
- “Replica Exchange” (RE) ;

**L’ensemble ’Nombre d’actifs limité’** L’ensemble “Nombre d’actifs limité” est composé de six groupes de tests avec un nombre de positions actives fixe :

1. aucune position active
2. une position active
3. cinq positions
4. dix positions
5. vingt positions
6. trente positions

### **3.1. Les méthodes pratiques**

---

Lorsqu'une position n'est pas active, on fixe l'acide aminé de la position en utilisant l'acide aminé de la séquence native.

Le groupe ‘aucune position active’ est constitué d'un test par algorithme pour chaque protéine. Il y a donc neuf tests par algorithme. Ce sont les tests pendant lesquels la séquence d'acides aminés est fixe et correspond à la séquence native de la protéine.

Pour les tests avec une seule position active, comme des temps de calcul le permettent, nous décidons d'être exhaustif : Toutes les positions sont testées, Il y a alors huit cent quatre tests par algorithme. Pour tous les autres groupes de tests (5,10,20 et 30 positions actives) cinq tests sont effectués par protéine, c'est-à-dire quarante-cinq tests par algorithme.

**le choix des positions actives** Pour définir complètement les tests ,il reste maintenant à décrire comment le choix des positions actives pour les groupes numéro trois à numéro six a été effectué. Il y a peu d'intérêt à tester des situations avec position active sans interaction avec les autres positions actives. En effet s'il existe une position active P dont chaque résidu est sans interaction avec tous les résidus possibles des autres positions actives, déterminer le meilleur état pour P est proche du test du groupe 2 avec P comme position active. Notons qu'en même que cela n'est pas exactement la même question parce que les positions actives différentes de P peuvent influencer la position de la chaîne latérale de positions inactives qui à leur tour peuvent influencer l'état de P. Ainsi le choix des positions actives se fait non pas pas tirage aléatoire, car le risque d'avoir les positions peu en inter action est trop grand. Il se fait sous contrainte d'interaction.

**positions en interactions** Pour cela nous utilisons la notion de voisinage incluse dans proteus : Deux positions P et Q sont en interaction s'il existe un rotamère  $r_P$  de P et un rotamère  $r_Q$  de Q tels que :

$$|E(r_P, r_Q)| > S_{Vois}$$

avec  $S_{Vois}$  un seuil donné par utilisateur à la configuration de proteus (voir chap. ?? pour les détails).

Définissons maintenant la notion de n-uplet en interaction par la donnée de n positions avec  $n \in \{5,10,20,30\}$  et d'un seuil  $S_{Vois}$  tels que pour toute paire de positions (P,Q) du n-uplet, P et Q sont en interaction.

Pour déterminer les positions actives,nous exécutons proteus en mode verbeux sans effectué d'optimisation. Il existe plusieurs façons de procéder ,ici nous utilisons le mode Monte-Carlo avec une trajectoire de zéro pas. Ces exécutions produisent en sortie standard

### **Chapitre 3. Les comparaisons d’algorithmes**

---

la liste des voisins pour chaque position au seuil donnée en paramètre. Pour chaque protéine, nous exécutons protéus avec  $S_{Vois}$  égal à 10 , 5 , 1 et 0.1 à tour de rôle. Nous obtenons quatre listes de voisins. Ensuite, nous cherchons dans les listes, par un script dédié, les n-uplets en interaction en partant de la liste de voisins au sens le plus fort, c'est-à-dire 10, vers la liste de voisins au sens le plus faible (0.1). La recherche s'arrête si cinq n-uplets sont trouvés.

#### **Définition de protocole comparable**

Nous voulons comparer les algorithmes très différents. Un algorithme peut garantir l'obtention du minimum global en énergie (GMEC) s'il termine, mais ne garantit pas qu'il termine, un autre permet un contrôle très fin du temps d'exécution sans garantie du GMEC, et d'autres enfin ont des objectifs plus ambitieux qu'uniquement l'obtention du GMEC. Mais le GMEC reste le meilleur point de commun. Nous allons donc y concentrer les comparaisons.

Nous devons noter également que l'obtention du GMEC est théorique, en pratique nous n'avons pas de preuve que le code de l'algorithme exact que nous utilisons soit sans bogue. Cependant, nous mettons de côté cette éventualité et dans toute la suite GMEC désigne aussi bien le minimum global en énergie que le résultat de toulbar2 lorsqu'il termine.

Le Monte-Carlo et le “Replica exchange” possèdent de nombreux paramètres de configuration, ce qui rend l'ensemble des protocoles possibles très grand. Se pose alors la question de l'optimisation du protocole. L'objectif que nous nous fixons ici, n'est pas de rechercher un protocole optimum pour chacun des tests, mais de tester un protocole optimisé par algorithme. Nous allons alors dans un premier temps, rechercher les meilleurs protocoles Monte-Carlo et les meilleurs protocoles “Replica Exchange” sur l'ensemble de tests “tout actifs”. Puis, sur la base des résultats obtenus, nous ferons l'ensemble des tests à positions actives limitées avec les meilleurs protocoles. Le programme toulbar2 possède aussi de nombreuses options. Deux paramétrages différents seront utilisés.

Pour rendre les protocoles comparables, le temps d'exécution maximum est fixé à vingt-quatre heures, sauf mention contraire. Toulbar2 donne sa meilleure séquence/conformation en dernier, il n'y a donc pas post-traitement nécessaire. C'est également le cas pour le MC à condition de configurer l'impression de la trajectoire avec la balise *Print\_Threshold = 0*. dans le fichier de configuration. Pour le RE et l'Heuristique, un tri des séquences selon l'énergie est nécessaire. Mais il n'y a pas beaucoup de séquences :

1. L' Heuristique fournit une séquence/conformation à chaque cycle.

2. Le RE avec  $Print\_Threshold = 0$  produit autant de fichiers de séquences/rotamères que de marcheurs. Chacun ne contenant pas plus de quelques dizaines de séquences/rotamères.

Nous pouvons donc négliger la durée du tri dans le temps total d'exécution.

#### Protocole Heuristique

Pour l'algorithme Heuristique, il n'y a dans notre situation qu'un seul paramètre à renseigner : le nombre de cycles à effectuer. Quelques essais préliminaires sur la plus grosse protéine (Table 3.1) dans le cas où tout actif , montre que la version utilisée de proteus peut effectuer jusqu'à environ 110000 cycles sur nos machines de calculs en l'espace de vingt-quatre heures. Ainsi le protocole H est défini comme le protocole qui utilise le mode heuristique de proteus et qui effectue cent dix mille cycles. Nous définissons également les variantes H-, H+ et H++ comme des protocoles plus courts ou plus longs à facteur entier près (Table 3.2). Par ailleurs certains comparaisons de l'heuristique avec la MC ont été faites avec une version précédente de proteus notons h ce protocole. Il diffère aussi de H par le fait que les options d'optimisation du compilateur Intel -O2 contre -O3 pour H.

Nom	nombre de cycles
H	110000
H-	1100
H+	330000
H++	990000
h	100000

Table 3.2 – Les protocoles Heuristique

#### 3.1.1 Protocoles Monte-Carlo

On distingue deux ensembles de protocoles, Le premier où les noms sont de la forme mc\*, rassemble les protocoles utilisés pendant l'optimisation du Monte-Carlo. Le second est constitué du protocole retenu comme optimisé , plus une variante.

#### Optimisation

Les éléments à paramétrier pour l'algorithme Monte-Carlo sont les suivants :

1. la température
2. le nombre de pas (avec le nombre de trajectoires et la longueur de trajectoire )

### **Chapitre 3. Les comparaisons d'algorithmes**

---

3. Le seuil de voisinage
4. Les probabilités de changements de la séquences/rotamère

Ce qui représente un ensemble de protocoles trop grand pour une approche exhaustive. Nous allons faire varier les paramètres un par un.

La température est le paramètre principale du Monte-carlo, c'est lui qui contrôle le taux d'acceptation du critère de Metropolis. Alors la première étape de cette optimisation va consisté à faire variés la température , entre 0.001 et 0.5 , en conservant les autres paramètres fixés (protocoles de mc0 à mc5). Le nombre de pas total effectué est le produit de deux paramètres, le nombre de trajectoires et la longueur de trajectoire. les protocoles mc1b et mc2b testent l'effet d'une augmentation du nombre de pas. Tandis que mc2d teste l'effet de variation du rapport nombre de trajectoire sur la longueur. Le protocole mc2e s'intéresse aux probabilités de changement de la trajectoire. Il y a cinq balises dans proteus qui contrôle ces changements :

1. <Prot> donne la probabilités de modifications de rotamère à une postions.
2. <Prot\_Prot> donne la probabilités de modifications de rotamère à deux positions.
3. <Mut> donne la probabilités de modifications de type de résidu à une position.
4. <Mut\_Prot> donne la probabilités de modifications de rotamères à deux positions.
5. <Mut\_Mut> donne la probabilités de modifications de type de résidu à deux positions.

La table 3.3 donne les probabilités utilisées par ces cinq paramètres dans l'ordre de la liste précédente.

Enfin mc4b se distingue des autres par un seuil de voisinage plus grand ((Table 3.3)).

Pour l'étape suivante, qui consiste à la comparaison avec les autres algorithmes, deux protocoles sont utilisés MC2d2 et MC2e2 ils inspirent de mc2d et mc2e , en étant adapté à la contrainte du temps de calculs de la comparaison d'algorithmes et un utilisant la nouvelle version de proteus.

#### **3.1.2 Les protocoles Replica Exchange**

L'algorithme “Replica Exchange” (RE) est une extension du Monte-Carlo , Les paramètres d'un protocole RE sont ceux d'un protocole Monte-Carlo plus d'autres paramètres :

1. le nombre de marcheurs
2. la température pour chaque marcheur

### 3.1. Les méthodes pratiques

Nom	Temp	Long. de trajectoire(mega)	Nb de trajectoires	Voisin	Proba
mc0	0.001	3	1000	10	0 ; 1 ; 0.1 ; 0 ;0
mc1	0.1	3	1000	10	0 ; 1 ; 0.1 ; 0 ;0
mc2	0.2	3	1000	10	0 ; 1 ; 0.1 ; 0 ;0
mc3	0.3	3	1000	10	0 ; 1 ; 0.1 ; 0 ;0
mc4	0.5	3	1000	10	0 ; 1 ; 0.1 ; 0 ;0
mc5	0.7	3	1000	10	0 ; 1 ; 0.1 ; 0 ;0
mc1b	0.1	6	1000	10	1 ; 1 ; 1 ; 1 ;0
mc2b	0.2	6	1000	10	0 ; 1 ; 0.1 ; 0 ;0
mc2c	0.2	3	10000	10	0 ; 1 ; 0.1 ; 0 ;0
mc2d	0.2	3000	1	10	0 ; 1 ; 0.1 ; 0 ;0
mc2e	0.2	3	1000	10	1 ; 0 ; 0.1 ; 0 ;0
mc4b	0.5	10	100	20	0 ; 1 ; 0 ; 1 ;0
MC2d2	0.2	6000	1	10	0 ; 1 ; 0.1 ; 0 ;0
MC2e2	0.2	6000	1	10	1 ; 0 ; 0.1 ; 0 ;0

Table 3.3 – Les protocoles Monte-Carlo

3. la période de “swap”, c'est-à-dire la période (en nombre de pas) à laquelle le test de Hasting sur l'échange de température et éventuellement l'échange ,sont effectués.

Pour avoir des exécutions en parallèle sans plusieurs marcheurs par cœur du processeur, nous limiter nos tests à quatre ou huit marcheurs. La distribution des températures est un éléments déterminant dans le comportement des marcheurs, car c'est elle qui pilote en grande partie le taux d'acceptation des échanges de températures. Nous suivant l'idée proposée par Kofke de lui faire suivre une progression géométrique (  $\frac{T_i}{T_{i+1}} = C$  , avec C une constante) [Kofke 2002, 2004a,b]. Ceci garantie alors que le taux d'acceptation d'échange entre  $T_i$  et  $T_{i+1}$  soit égale pour tout nos i. De plus nous souhaitons centrer à peu près, nos distributions sur la température ambiante.

Voici les températures pour le RE quatre marcheurs :

1. 10, 1, 0.1 et 0.01
2. 2, 1, 0.5 et 0.25
3. 1, 0.5, 0.25 et 0.125

,et celles pour le RE huit marcheurs :

1. 3 , 2 , 1.333 , 0.888 , 0.592 , 0.395 , 0.263 et 0.175
2. 10 , 3.16 , 1 , 0.316 , 0.1 , 0.0316 , 0.01 et 0.00316

Ici les protocoles ne se font qu'avec une seule trajectoire par marcheur. Et la contrainte du temps de calculs se comprends comme vingt-quatre heures de calculs cumulées sur tous

### ***Chapitre 3. Les comparaisons d’algorithmes***

---

les marcheurs. Ainsi les longueurs de trajectoire sont définies pour le RE quatre marcheurs comme le quart d’une trajectoire MC, pour le RE huit marcheurs comme le huitième.

La table 3.4 donne les probabilités utilisées par les cinq balises qui contrôlent les modifications de la séquence/conformation à chaque pas, dans l’ordre de la liste de la section 3.1.1.

Nom	marcheurs	Temp	Traj (mega)	seuil voisin	Proba	swap period (mega)
RE4a	4	10<->0.01	1500	10	1 ; 0 ; 0.1 ; 0 ; 0	7.5
RE4b	4	1<->0.125	1500	10	1 ; 0 ; 0.1 ; 0 ; 0	7.5
RE4c	4	2<->0.25	1500	10	1 ; 0 ; 0.1 ; 0 ; 0	7.5
RE8a1	8	10<->0.00316	750	10	1 ; 0 ; 0.1 ; 0 ; 0	1
RE8a2	8	10<->0.00316	750	0	1 ; 0 ; 0.1 ; 0 ; 0	2.5
RE8b1	8	3<->0.175	750	10	1 ; 0 ; 0.1 ; 0 ; 0	7.5
RE8b2	8	3<->0.175	750	10	0 ; 1 ; 0.1 ; 0 ; 0	7.5
RE8b3	8	3<->0.175	750	10	0 ; 1 ; 0.1 ; 0 ; 0	1

Table 3.4 – Les protocoles Replica Exchange

#### **3.1.3 Les protocoles Toulbar2**

### 3.1.4 Les temps de calculs

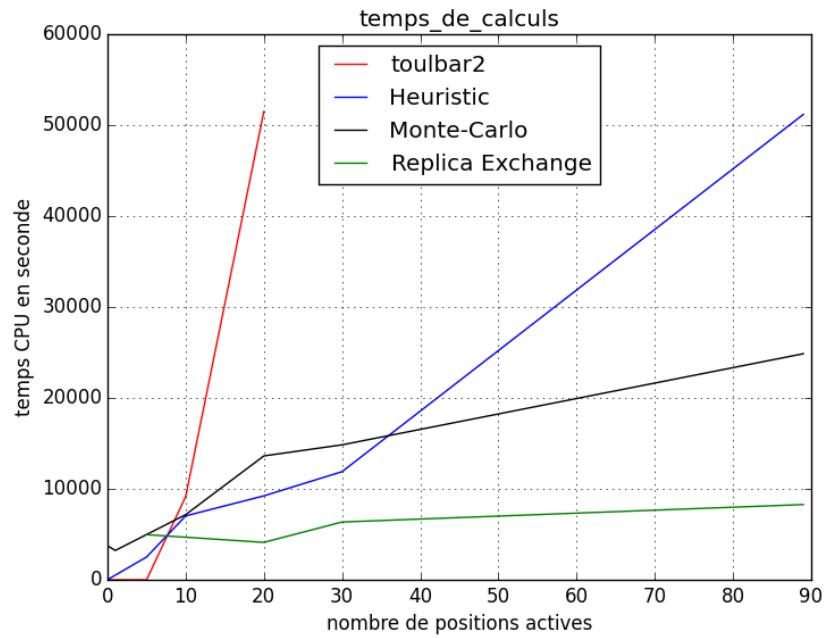


Figure 3.1 – Temps d’occupation du processeur selon le nombre de positions actives.

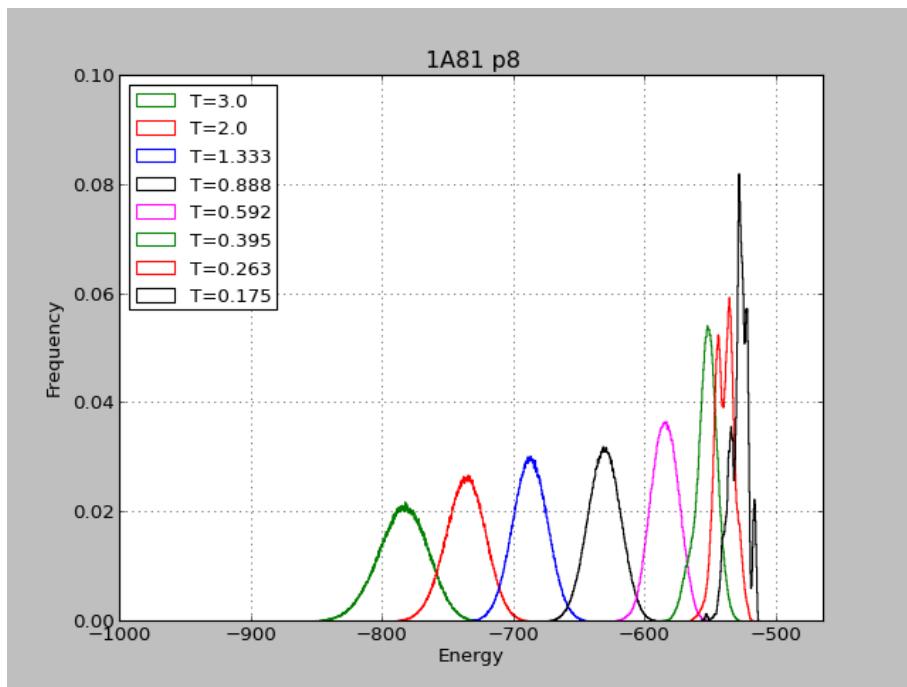


Figure 3.2 – Distribution des énergies selon la température (protocole RE3).

## 3.2 Les tests

### 3.2.1 Tous les résidus actifs

#### Les meilleures énergies

Protéine	h	MC3	MC43	RE1	RE2	RE5	RE3	RE32	RE4
1A81	-521	-538	-522	-525	-520	-520	-514	-512	-518
1ABO	-272	-274	-268	-273	-269	-273	-268	-271	-272
1BM2	-484	-500	-486	-488	-481	-489	-478	-476	-486
1CKA	-252	-258	-249	-259	-251	-251	-247	-246	-249
1G9O	-428	-435	-428	-429	-421	-430	-428	-425	-428
1M61	-480	-493	-479	-483	-480	-481	-480	-480	-480
1O4C	-535	-545	-531	-536	-529	-536	-527	-524	-532
1R6J	-407	-419	-414	-415	-409	-411	-409	-408	-414
2BYG	-457	-469	-454	-461	-456	-460	-456	-454	-462

Table 3.5 – les meilleures énergies pour tous les résidus actifs

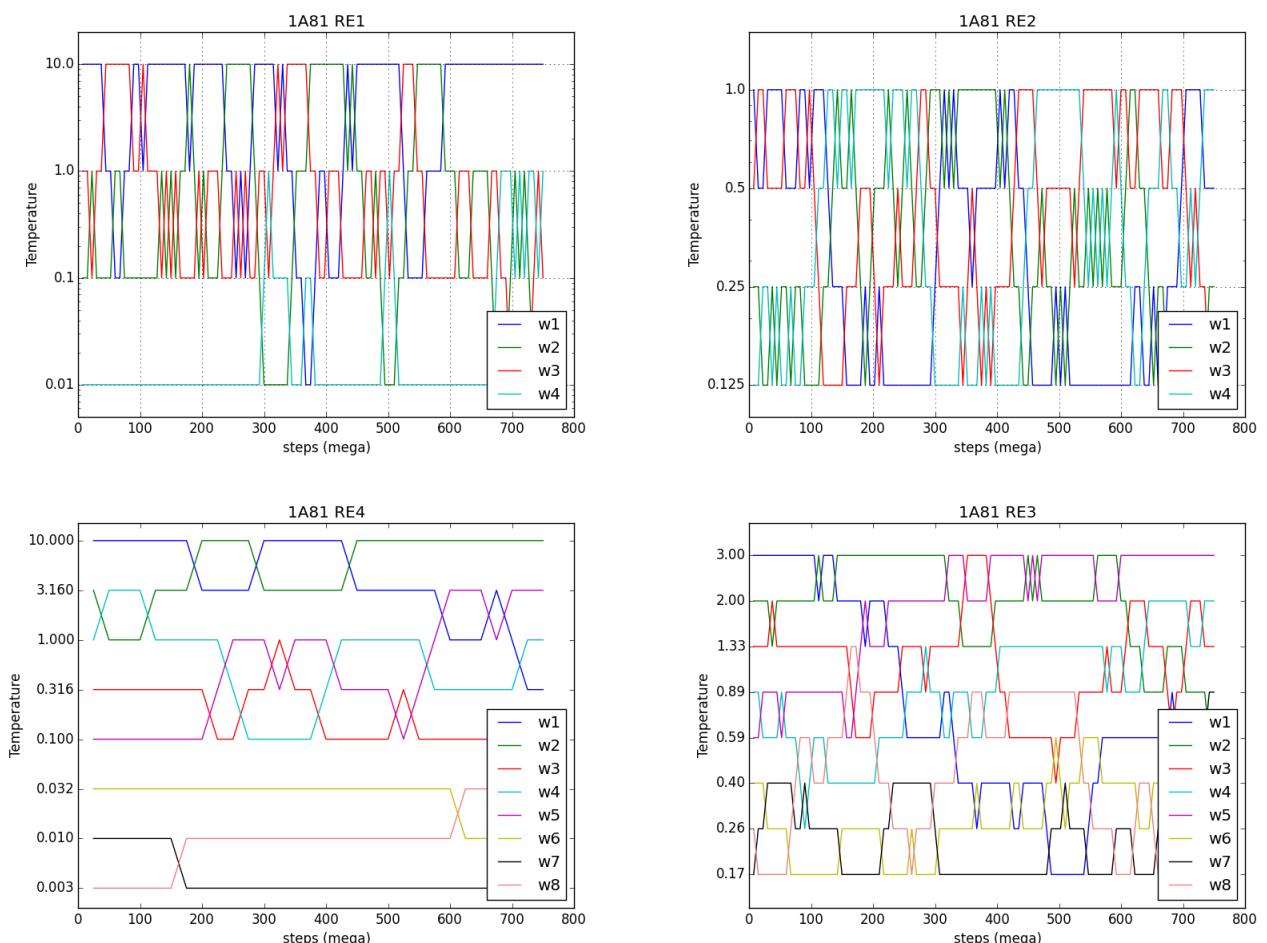


Figure 3.3 – Variation de la température au court de la trajectoire de chaque marcheur (protocole RE1).

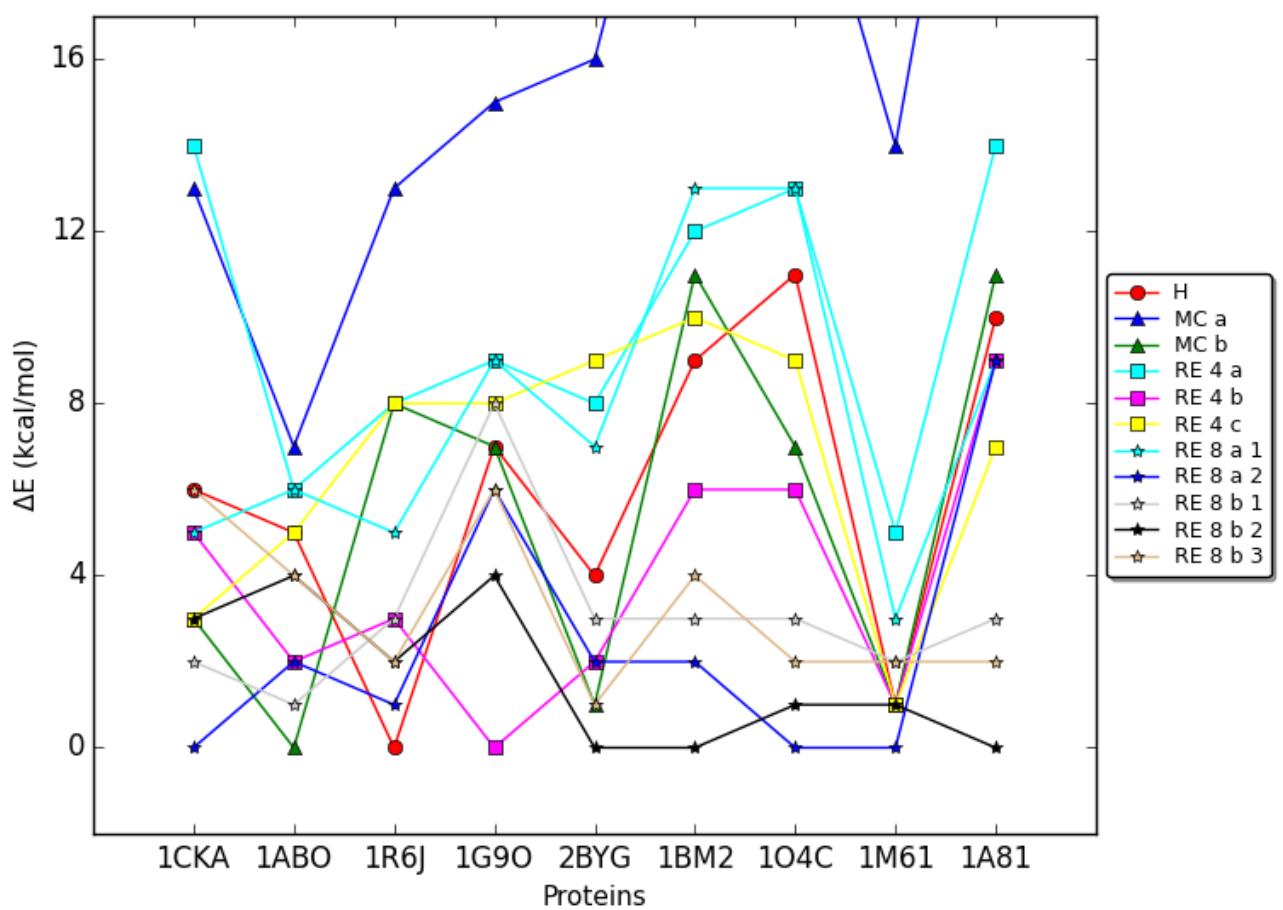


Figure 3.4 – Tous les protocoles.

### 3.2. Les tests

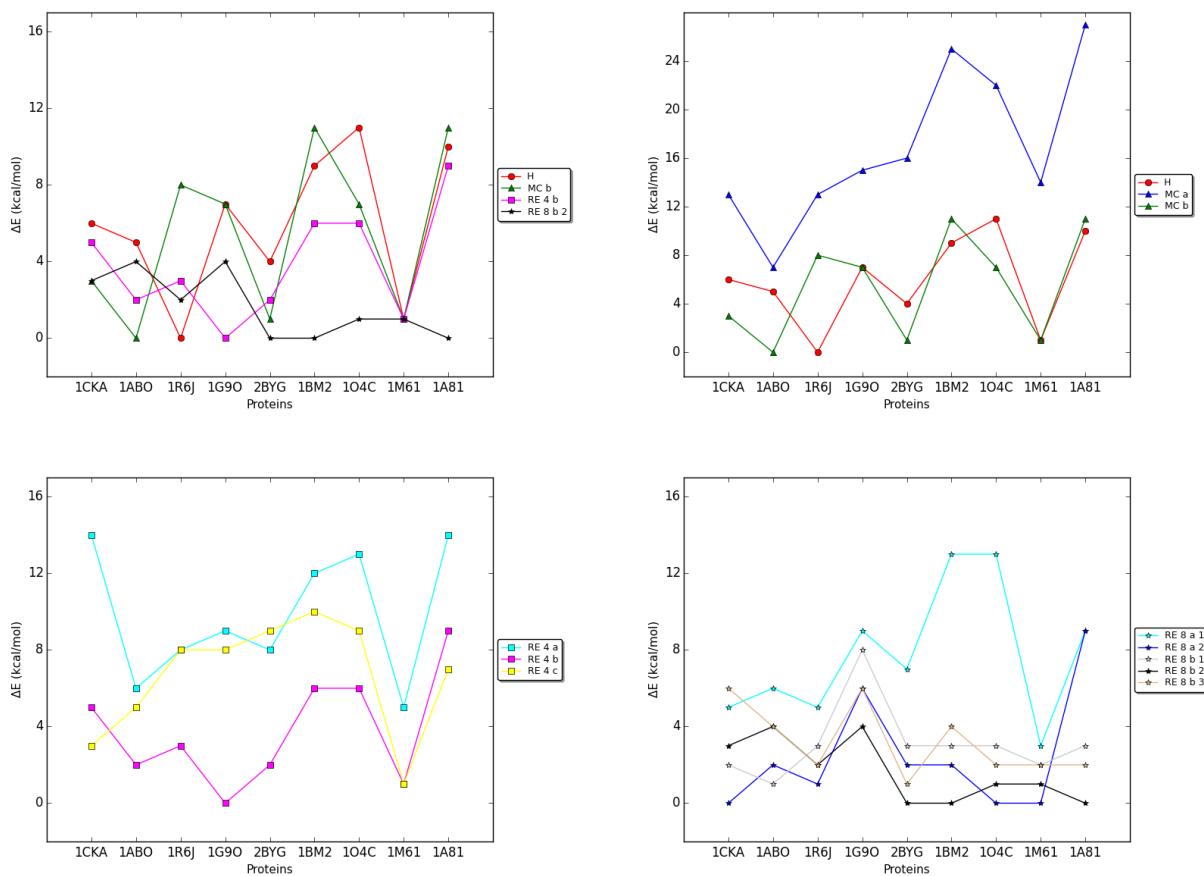


Figure 3.5 – Variation de la température au court de la trajectoire de chaque marcheur (protocole RE1).

### 3.2.2 Tous les résidus inactifs

#### Séquence native

Protéine	GMEC	H-	MC0	MC4-
1A81	-585.1365	0	-0.2547	0
1ABO	-320.1798	0	0	0
1BM2	-553.5532	0	-0.0564	-0.0121
1CKA	-319.2787	0	0	0
1G9O	-481.1175	0	-0.1394	0
1M61	-555.9140	0	0	0
1O4C	-591.2115	0	0	-0.1250
1R6J	-454.9340	0	0	0
2BYG	-507.0165	0	0	0

Table 3.6 – L’énergie du GMEC et la différence avec les autres protocoles. Tous les résidus inactifs ()

#### Une position active

Position	GMEC	MC4-
14	-584.4693	-0.0405
39	-584.7378	-0.0111
55	-584.0477	-0.0012
60	-583.7763	-0.0140
66	-592.3835	-0.0347
70	-583.8950	-0.0348
71	-588.5916	-0.0247
76	-583.3815	-0.0248
79	-582.8485	-0.0406
86	-584.1412	-0.0248
101	-583.8406	-0.0248
105	-583.0197	-0.0248
107	-582.2241	-0.0248

Table 3.7 – Liste des échecs pour 1A81

Cinq positions actives

Dix positions actives

Dix positions actives, mutations

Vingt et trente positions actives

### **3.2.3 Etude au voisinage de GMECs**

Position	GMEC	MC4-
2	-553.3134	-0.0040
3	-553.5532	-0.0121
5	-553.0932	-0.0179
6	-553.5532	-0.0121
8	-556.1917	-0.0148
10	-551.4990	-0.0149
11	-551.8859	-0.0149
12	-550.8152	-0.0148
13	-553.4829	-0.0451
14	-553.5532	-0.0121
15	-553.5532	-0.0121
17	-553.5532	-0.0121
18	-553.0880	-0.0121
19	-553.5532	-0.0270
20	-553.0003	-0.0121
21	-553.5532	-0.0121
22	-553.1769	-0.0121
29	-553.5532	-0.0121
34	-553.5532	-0.0270
36	-555.3358	-0.0317
37	-553.5532	-0.0121
41	-553.5076	-0.0121
46	-552.9056	-0.0149
49	-553.5532	-0.0121
51	-553.5532	-0.0179
55	-551.8384	-0.0121
56	-553.5532	-0.0121
57	-561.0695	-0.0121
58	-553.5532	-0.0121
62	-553.5532	-0.0121
65	-553.5532	-0.0121
66	-551.2026	-0.0179
68	-552.6182	-0.0148
70	-553.5532	-0.0121
72	-552.2724	-0.0121
73	-553.5532	-0.0121
75	-553.5532	-0.0179
77	-553.0234	-0.0466
80	-553.5532	-0.0121
81	-553.5532	-0.0121
82	-548.0641	-0.0121
83	-553.5532	-0.0121
85	-550.1884	-0.0122
86	-552.7375	-0.0148
87	-550.6139	-0.0121
90	-552.8601	-0.0009
91	-553.5532	-0.0121
92	-553.5532	-0.0121
93	-553.2772	-0.0148
94	-553.3207	-0.0251

Position	GMEC	MC4-
17	-316.1693	-0.0109

Table 3.9 – Liste des échecs pour 1CKA

Position	GMEC	MC4
58	-561.9469	-0.0138

Table 3.10 – Liste des échecs pour 1M61

Position	GMEC	MC4-
1	-591.2115	-0.1380
2	-591.2115	-0.1250
3	-591.2115	-0.1250
4	-590.7216	-0.0319
5	-590.5458	-0.1071
6	-591.2115	-0.1521
7	-590.7923	-0.1429
8	-591.2115	-0.1250
9	-591.2115	-0.1728
10	-591.2115	-0.2572
11	-589.9443	-0.2489
12	-591.1022	-0.1137
13	-589.9867	-0.0535
14	-591.2115	-0.1250
15	-589.4899	-0.0436
16	-591.2115	-0.1521
17	-590.4460	-0.0557
18	-589.0053	-0.1366
19	-590.7580	-0.0348
20	-591.2115	-0.1250
21	-591.2115	-0.1600
22	-591.2115	-0.1250
23	-590.5249	-0.1530
24	-590.7262	-0.0630
25	-591.2115	-0.1250
26	-591.2115	-0.1250
27	-590.8058	-0.1194
28	-591.2115	-0.1250
29	-591.2115	-0.1571
30	-590.5207	-0.0221
31	-590.5507	-0.0530
32	-591.2115	-0.1571
33	-591.2115	-0.1234
34	-590.7486	-0.1258
35	-591.2115	-0.0378
36	-589.1510	-0.0974
37	-591.0133	-0.0941
38	-589.2126	-0.2743
39	-589.0387	-0.1890
40	-590.8793	-0.0883
41	-589.4209	-0.0409
42	-591.2115	-0.1250
43	-587.9420	-0.1315
44	-589.8470	-0.0595
45	-591.2115	-0.1712
46	-588.8346	-0.2668
47	-589.9117	-0.2773
48	-588.6520	-0.2625
49	-591.2115	-0.2120
50	-590.6561	-0.0807

Position	GMEC	MC4-
4	-453.4484	-0.0155
20	-452.6464	-0.0114
32	-454.9340	-0.0092
68	-454.4856	-0.0060
73	-454.7809	-0.0155
77	-454.1344	-0.0155
79	-453.4729	-0.0155

Table 3.12 – Liste des échecs pour 1R6J

Position	GMEC	MC4-
1	-505.2910	-0.0132
3	-506.7960	-0.0254
4	-505.5800	-0.0023
5	-506.8732	-0.0948
49	-505.5183	-0.0135
59	-507.0165	-0.0100
85	-506.6217	-0.0101
88	-505.2286	-0.0097
95	-506.3195	-0.0131

Table 3.13 – Liste des échecs pour 2BYG

### Chapitre 3. Les comparaisons d’algorithmes

---

Protéine	GMEC	H	MC4	RE3
1A81 1	-579.3989	0	0	
1A81 2	-575.2254	0	0	
1A81 3	-582.7452	0	0	
1A81 4	-569.9383	0	-5.3443	0
1A81 5	-591.8143	0	0	
1ABO 1	-315.4497	0	0	
1ABO 2	-316.6637	0	0	
1ABO 3	-307.4824	0	0	
1ABO 4	-313.7710	0	0	
1ABO 5	-313.5695	0	0	
1BM2 1	-548.2341	0	0	
1BM2 2	-554.8135	0	0	
1BM2 3	-557.8629	0	0	
1BM2 4	-544.9791	0	0	
1BM2 5	-550.2956	0	-0.0121	
1CKA 1	-315.0859	0	0	
1CKA 2	-309.7692	0	0	
1CKA 3	-317.3820	0	0	
1CKA 4	-314.8550	0	0	
1CKA 5	-312.0405	-0.0001	-0.0001	
1G9O 1	-469.9540	0	0	
1G9O 2	-476.4094	0	0	
1G9O 3	-479.7190	0	0	
1G9O 4	-478.9513	0	0	
1G9O 5	-480.7260	0	0	
1M61 1	-557.6647	0	0	
1M61 2	-546.9587	0	0	
1M61 3	-553.0731	0	0	
1M61 4	-555.0885	0	0	
1M61 5	-554.6356	0	0	
1O4C 1	-584.4267	0	-0.0655	
1O4C 2	-584.8989	0	-0.1437	
1O4C 3	-588.4971	0	-0.1164	
1O4C 4	-587.7129	0	-0.1400	
1O4C 5	-587.6514	0	-0.1168	
1R6J 1	-444.5018	0	0	
1R6J 2	-449.3043	0	-0.9421	0
1R6J 3	-453.1139	0	0	
1R6J 4	-453.1139	0	0	
1R6J 5	-454.9340	0	0	
2BYG 1	-500.7946	0	-0.0150	
2BYG 2	-506.2319	0	0	
2BYG 3	-506.8744	0	-0.0131	
2BYG 4	-504.5135	0	0	
2BYG 5	-506.0052	0	0	

### 3.2. Les tests

Protéine	GMEC	toulbar2	H	MC	RE
1A81 1	yes	-583.9354	0.	0.	
1A81 2	yes	-581.7802	0.	0.	
1A81 3	yes	-587.4392	-0.0001	-0.1595	
1A81 4	yes	-589.1322	0.	-0.0317	
1A81 5	yes	-578.2558	0.	-0.0563	
1ABO 1	yes	-309.1670	-0.0675	-0.9054	
1ABO 2	yes	-308.8387	0.	0.	
1ABO 3	yes	-303.8520	0.	0.	
1ABO 4	yes	-310.0087	0.	-0.0128	
1ABO 5	yes	-301.6727	0.	0.	
1BM2 1	yes	-549.8638	0.	-0.0950.	
1BM2 2	yes	-541.5944	0.	0.	
1BM2 3	yes	-543.7434	0.	0.	
1BM2 4	yes	-549.0453	0.	0.	
1BM2 5	yes	-544.1447	0.	-0.1082	
1CKA 1	yes	-305.8477	0.	0.	
1CKA 2	yes	-309.9886	0.	0.	
1CKA 3	yes	-304.6618	0.	0.	
1CKA 4	yes	-302.4894	0.	0.	
1CKA 5	yes	-299.2329	-0.2859	-3.2525	0.
1G9O 1	yes	-466.6764	0.	0.	
1G9O 2	yes	-478.8797	0.	0.	
1G9O 3	yes	-477.2503	-0.1366	0.	
1G9O 4	yes	-470.6458	0.	0.	
1G9O 5	yes	-464.8659	0.	-3.9599	0.
1M61 1	yes	-550.0699	0.	-0.0776	
1M61 2	yes	-538.6026	-3.5105	-4.5062	0.3215
1M61 3	yes	-552.2673	0.	0.	
1M61 4	yes	-550.0553	0.	0.	
1M61 5	yes	-553.6559	0.	-0.0432	
1O4C 1	yes	-587.4665	0.	-0.1121	
1O4C 2	yes	-585.8545	0.	-0.1046	
1O4C 3	yes	-580.3505	0.	-0.1519	
1O4C 4	yes	-587.1548	0.	-0.1545	
1O4C 5	yes	-590.2650	0.	-0.1753	
1R6J 1	yes	-448.8351	0.	-2.4022	-2.3986
1R6J 2	yes	-448.4631	0.	-1.0398	
1R6J 3	yes	-450.3950	0.	-0.0106	
1R6J 4	yes	-451.7211	0.	0.	
1R6J 5	yes	-450.9943	0.	-0.0162	
2BYG 1	no	-5.7485	-505.6397	-0.0337	
2BYG 2	yes	-504.7389	0.	0.	
2BYG 3	yes	-504.3048	0.	-0.0833	
2BYG 4	yes	-504.3466	0.	-0.2149	
2BYG 5	yes	-491.6095	0.	0.	

Table 3.15 – Résultats 10 positions actives

Protéine	H mut nb	MC mut nb
1A81 1	0	0
1A81 2	0	0
1A81 3	0	2
1A81 4	0	0
1A81 5	0	0
1ABO 1	0	4
1ABO 2	0	0
1ABO 3	0	1
1ABO 4	2	2
1ABO 5	0	0
1BM2 1	0	2
1BM2 2	0	0
1BM2 3	0	0
1BM2 4	0	1
1BM2 5	0	2
1CKA 1	0	0
1CKA 2	0	1
1CKA 3	0	0
1CKA 4	0	0
1CKA 5	5	3
1G9O 1	0	0
1G9O 2	0	0
1G9O 3	0	0
1G9O 4	0	0
1G9O 5	0	3
1M61 1	0	2
1M61 2	3	7
1M61 3	0	0
1M61 4	0	0
1M61 5	0	0
1O4C 1	0	0
1O4C 2	0	0
1O4C 3	0	0
1O4C 4	0	0
1O4C 5	0	3
1R6J 1	0	3
1R6J 2	0	2
1R6J 3	0	0
1R6J 4	0	0
1R6J 5	0	0
2BYG 1	no	no
2BYG 2	0	0
2BYG 3	0	1
2BYG 4	1	3
2BYG 5	0	0

### 3.2. Les tests

---

Protéine	GMEC	toulbar2	H	MC	RE
1A81 1	yes	-566.9106	0.	-0.3275	-0.3851
1A81 2	yes	-564.6618	-0.1705	-2.4355	-1.0069
1A81 3	yes	-572.7774	0.	-0.4640	-0.6186
1A81 4	yes	-572.9780	-0.3878	-0.5748	-0.6991
1A81 5	yes	-572.7410	-0.0068	-0.5088	-0.1541
1ABO 1	yes	-299.6592	-0.1205	-1.1159	-0.2153
1ABO 2	no	-13.8563	-298.3854	0.	0.
1ABO 3	no	-1.2190	-298.3854	0.	0.
1ABO 4	no	-1.9940	-297.8545	-0.0076	0.
1ABO 5	no	-3.5418	-297.8009	-0.9483	-0.9483
1BM2 1	yes	-526.0936	0.	-0.0619	-0.1584
1BM2 2	no	-7.5304	-525.3588	-0.0725	-0.0143
1BM2 3	yes	-534.3861	-0.0229	-0.4762	-0.2897
1BM2 4	no	-0.1186	-526.8307	-2.5883	-0.0789
1BM2 5	yes	-535.3334	-0.2396	-0.3746	-0.3746
1CKA 1	yes	-295.8571	0.	0.	0.
1CKA 2	yes	-295.3571	0.	0.	0.
1CKA 3	yes	-293.8687	0.	0.	0.
1CKA 4	no	-4.3122	-293.8687	0.	0.
1CKA 5	no	-4.2849	-293.4203	0.	0.
1G9O 1	no	-2.0574	-451.4604	-1.2525	-1.2525
1G9O 2	no	-3.2106	-453.2474	-0.2177	-0.1915
1G9O 3	no	-1.9008	-453.7856	-0.4417	-0.1019
1G9O 4	no	-0.5030	-456.7331	-0.3855	-0.1455
1G9O 5	no	-0.4298	-456.9981	-0.1495	-0.5114
1M61 1	yes	-528.0700	0.	0.	0.
1M61 2	yes	-528.7653	0.	0.	0.
1M61 3	yes	-530.0684	0.	0.	0.
1M61 4	yes	-534.5248	0.	0.	0.
1M61 5	yes	-548.0096	0.	-0.2521	-0.1345
1O4C 1	no	-574.0047	-0.3465	-0.0690	-0.0587
1O4C 2	no	-6.4214	-574.8584	-0.1963	-0.3175
1O4C 3	yes	-573.6314	0.	-0.3461	-0.0997
1O4C 4	yes	-575.8667	0.	-0.3640	-0.1382
1O4C 5	no	-573.3479	0.	-0.1131	-0.2206
1R6J 1	yes	-440.7417	0.	-0.2604	-0.2002
1R6J 2	yes	-437.2537	0.	-0.0071	-0.0183
1R6J 3	yes	-439.4335	0.	-0.0537	-0.0732
1R6J 4	yes	-439.5988	0.	-0.0639	-0.0601
1R6J 5	yes	-438.0222	0.	-0.0735	-0.0244
2BYG 1	yes	-496.2991	0.	-3.1878	-0.0257
2BYG 2	yes	-494.8723	0.	-0.0524	-0.0831
2BYG 3	yes	-494.4390	0.	-1.3564	-0.0826
2BYG 4	yes	-495.9213	0.	-0.1968	-0.6022
2BYG 5	no	-1.8604	-497.5123	-0.0933	-0.0386

### Chapitre 3. Les comparaisons d’algorithmes

---

Protéine	GMEC	toulbar2	H	MC	RE
1A81 1	no	-1.2074	-562.9572	-0.6353	
1A81 2	no	-2.5520	-570.2620	-0.0578	
1A81 3	no	-43.5263	-562.9572	-2.4996	-1.2025
1A81 4	no	-5.1300	-559.6145	-0.0305	
1A81 5	no	-3.2417	-553.1077	-1.9586	-0.5791
1ABO 1	no	-44.5504	-296.5680	0.	
1ABO 2	no	-12.7303	-294.8500	0.	
1ABO 3	no	-9.3870	-295.2689	-0.2630	
1ABO 4	no	-10.7691	-296.5680	0.	
1ABO 5	no	-4.3907	-296.5680	0.	
1BM2 1	no	-22.5876	-556.1168	-1.7290	-1.6013
1BM2 2	no	-22.1386	-556.7539	-1.9856	-1.5876
1BM2 3	no	-22.5410	-556.1168	-1.9990	-1.1541
1BM2 4	no	-15.2639	-556.8507	-2.2127	-2.3854
1BM2 5	no	-15.9890	-556.3240	-2.83542	-1.1937
1CKA 1	no	-6.2700	-293.4203	0.	
1CKA 2	no	-2.0995	-293.4203	0.	
1CKA 3	no	-47.0217	-291.9243	0.	
1CKA 4	no	-44.0830	-293.4203	0.	
1CKA 5	no	-8.8608	-293.2709	0.	
1G9O 1	no	-2.0816	-449.0890	-1.5942	0.
1G9O 2	no	-0.3270	-452.6676	-0.3126	
1G9O 3	no	-17.7150	-450.0341	-1.5667	-1.5667
1G9O 4	no	-2.9758	-453.9682	-1.4284	-1.6202
1G9O 5	no	-445.8910	-1.6890	-7.6985	-2.3857
1M61 1	no	-14.4935	-0.0097	-523.9321	0.
1M61 2	no	-5.0899	-531.3717	-1.8749	-0.0083
1M61 3	no	-3.5795	-527.2659	-0.0154	
1M61 4	no	-16.1511	-530.2666	0.	
1M61 5	no	-23.0927	-522.5696	0.	
1O4C 1	no	-14.9064	-571.4882	-0.3435	
1O4C 2	no	-58.1558	-570.1458	-0.0795	
1O4C 3	no	-9.9221	-569.9777	-0.1789	
1O4C 4	no	-5.7790	-568.9839	-0.0423	
1O4C 5	no	-9.9221	-569.9777	-0.1789	
1R6J 1	yes	-435.4258	0.0	-0.0246	
1R6J 2	no	-14.9800	-435.0087	-0.0957	
1R6J 3	no	-439.8187	-439.8187	-0.0440	
1R6J 4	no	-435.0087	-0.0	-0.0957	
1R6J 5	no	-435.0970	-0.7036	-1.8823	-0.0781
2BYG 1	no	-17.9752	-492.6879	-0.1592	
2BYG 2	no	-0.3832	-492.3568	-0.1502	
2BYG 3	no	-0.1442	-492.6879	-0.1593	
2BYG 4	no	-492.6821	-0.0958	-0.0050	
2BYG 5	no	-0.5003	-492.1595	-0.6876	

### 3.2. Les tests

---

Protein	GMEC	H	MC	RE
1CKA 3	-304.6618	0	0	
1CKA 4	-302.4894	0	0	
1CKA 5	-299.2329	-0.2859	-3.2525	0
1G9O 3	-477.2503	-0.1366	0	
1G9O 4	-470.6458	0	0	
1G9O 5	-464.8659	0	-3.9599	0
1M61 1	-550.0699	0	-0.0776	
1M61 2	-538.6026	-3.5105	-4.5062	0.3215
1M61 5	-553.6559	0	-0.0432	

Protein	seq-rot nb	gmec+1	H rank	MC rank	seq nb	gmec+1	H mut nb	MC mut nb
1CKA 3	67669		1	1	227		0	0
1CKA 4	4649		1	1	498		0	0
1CKA 5	1388		78	?	77		0	2
1G9O 3	354559		23	1	63		1	0
1G9O 4	22639		1	1	381		0	0
1G9O 5	8658395		1	?	11		0	3
1M61 1	11199153		?	?	21		3	7
1M61 2	11199153		1	1	88		0	0
1M61 5	16417604		1	1	83		0	0

### Chapitre 3. Les comparaisons d'algorithmes

---

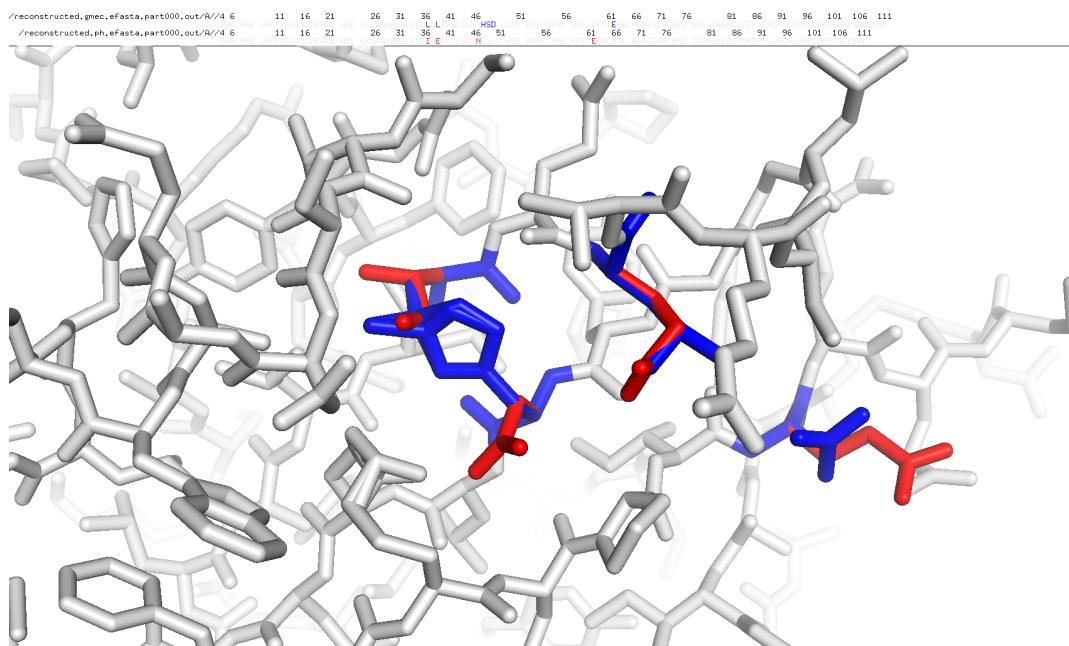


Figure 3.6 – test : 1M61 2, GMEC vs H

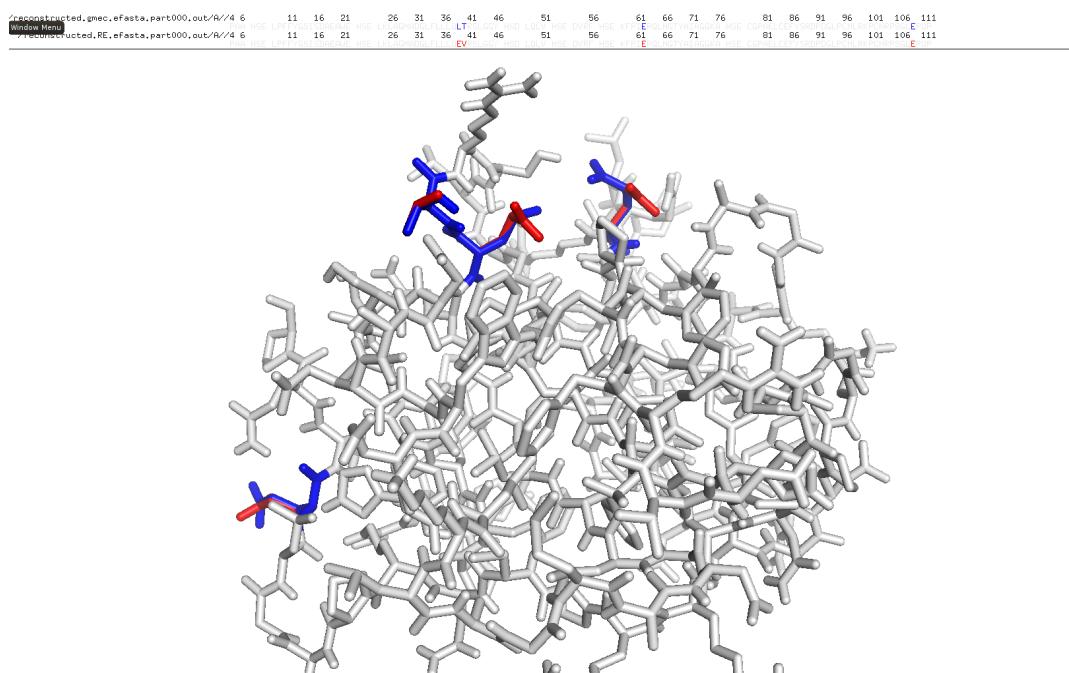


Figure 3.7 – test : 1M61 2, GMEC vs RE

### 3.2. Les tests

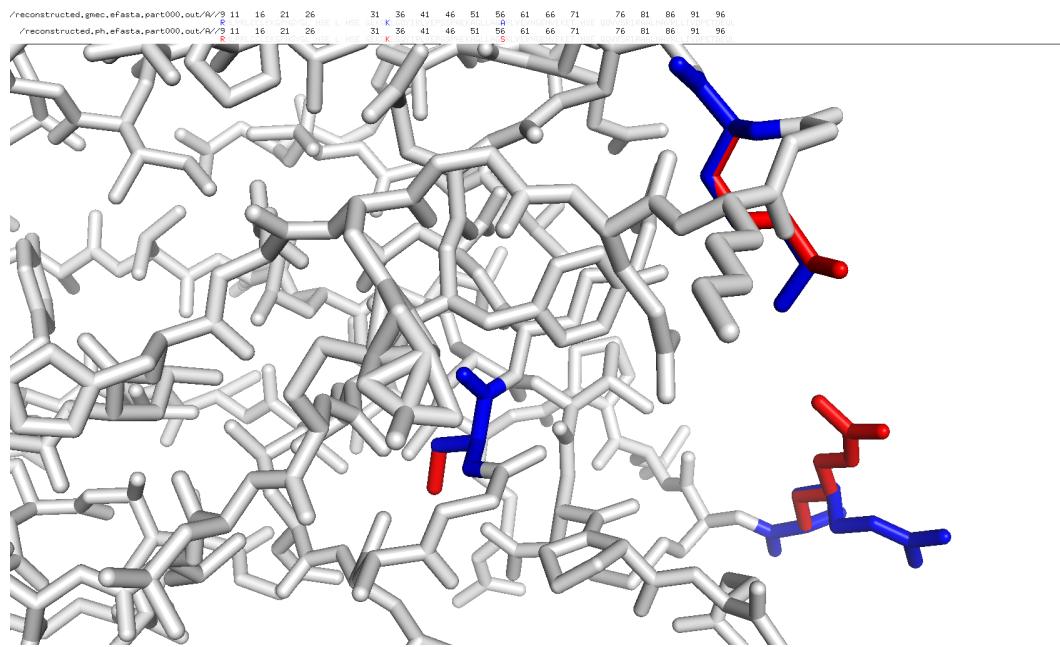


Figure 3.8 – test : 1G9O 3, GMEC vs H

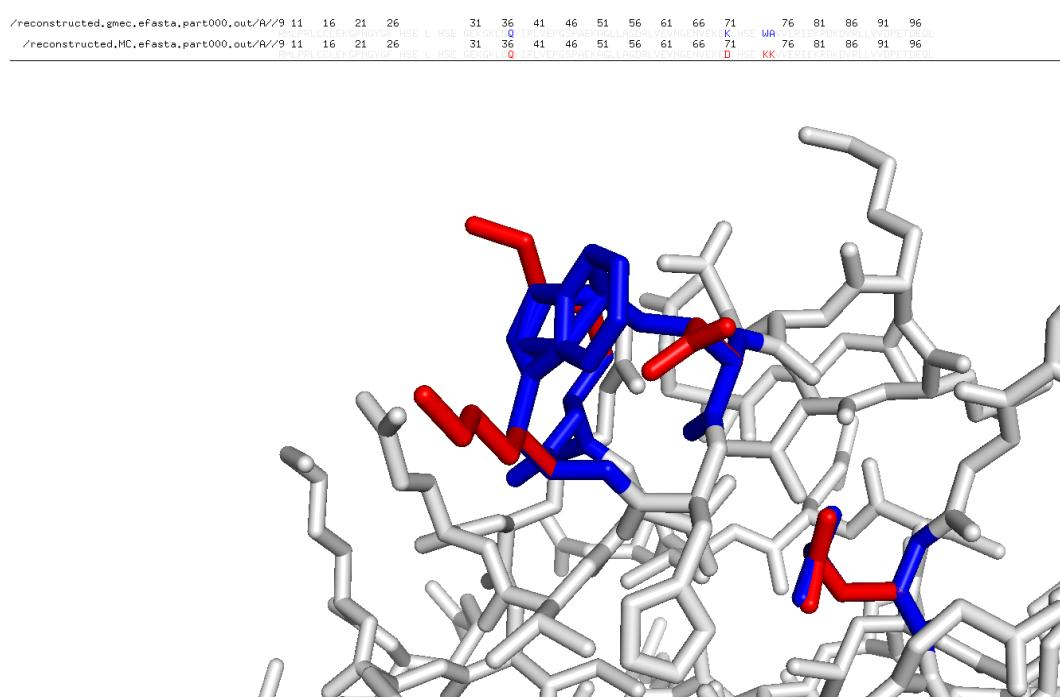


Figure 3.9 – test : 1G9O 5, GMEC vs MC

### Chapitre 3. Les comparaisons d'algorithmes

---

```
/reconstructed.gmec.eFasta.part000.out/A//134 141 146 151 156 161 166 171 176 181 186  
R D K E  
/reconstructed.ph.eFasta.part000.out/A//134 141 146 151 156 161 166 171 176 181 186  
R D K E
```

---

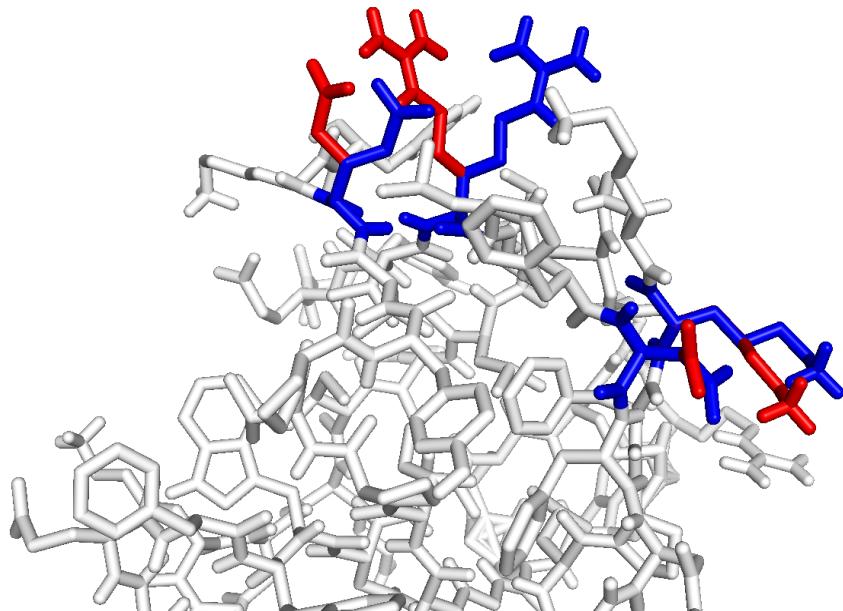


Figure 3.10 – test : 1CKA 5, GMEC vs H

```
/reconstructed.mc4.eFasta.part000.out/A//134 141 146 151 156 161 166 171 176 181 186  
/reconstructed.gmec.eFasta.part000.out/A//134 141 146 151 156 161 166 171 176 181 186  
R D K E  
R D G F
```

---

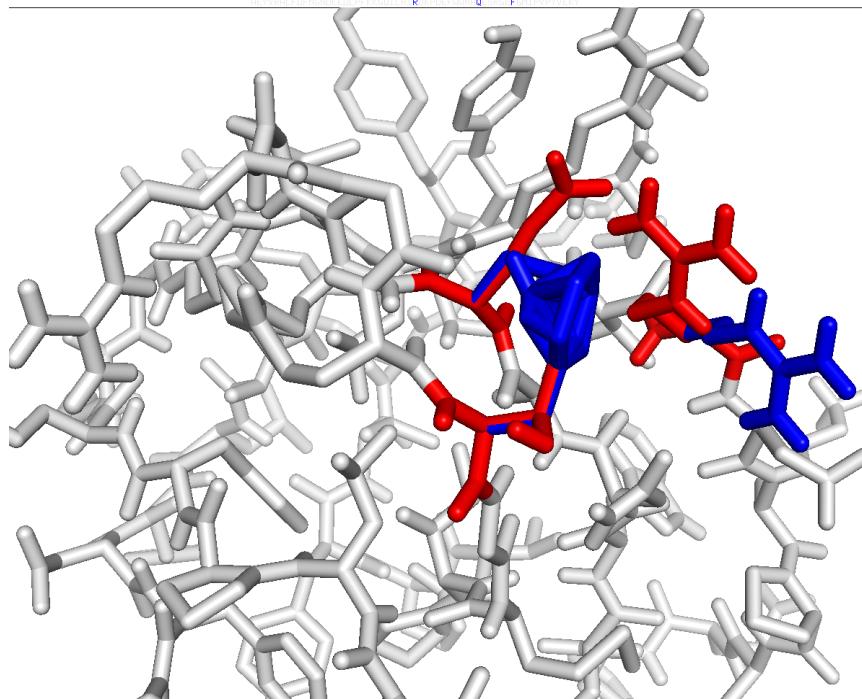


Figure 3.11 – test : 1CKA 5, GMEC vs MC

### 3.2. Les tests

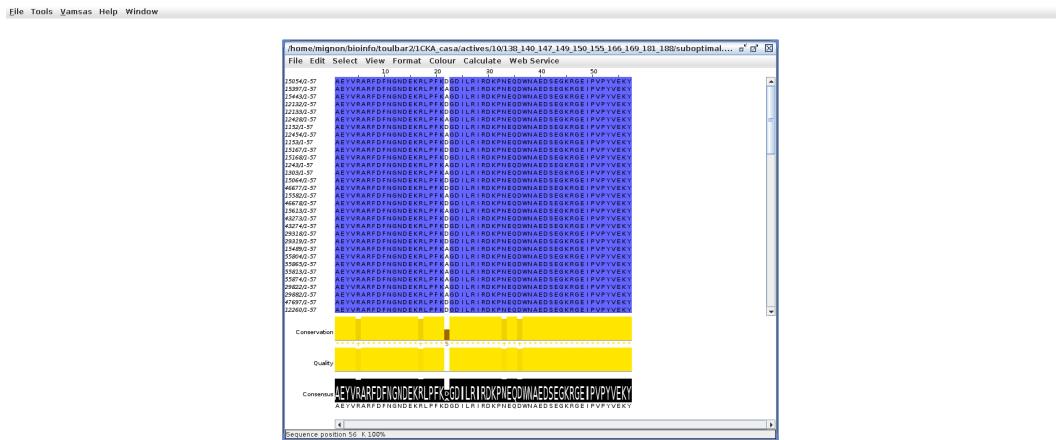


Figure 3.12 – Alignement du voisinnage : 1CKA\_5

#### 3.2.4 Résultats Superfamily

Protein	Match/seq size	Superfamily	Evalue	superfamily success	Family	Evalue	family success
1A81	no						
1ABO	51/58		4.4e-4	100%		2.8e-3	100%
1BM2	78/98		4.2e-5	100%		2.6e-3	100%
1CKA	40/57		1.1e-5	100%		3.4e-3.	100%
1G9O	79/91		7.0e-7	100%		2.5e-3	100%
1M61	97/109		7.2e-7	100%		2.6e-4	100%
1O4C	95/104		2.1e-4	100%		4.5e-3	100%
1R6J	74/82		9.8e-6	100%		4.6e-3	100%
2BYG	59/97		1.4e-5	100%		7.1e-3	100%

### 3.2.5 Résultats Heuristic (protocoles longs)

Proteins	GMEC	H	H+	H++
1ABO 1	-309.1670	-0.0675	-0.0675	0
1CKA 5	-299.2329	-0.2859	-0.0640	0
1G9O 3	-477.2503	-0.1366	0	0
1M61 2	-538.6026	-3.5105	-2.1673	-0.0188

Table 3.17 – Résultats pour 3 fois (resp 9 fois) plus de cycles heuristiques protocole H+ (resp H++)

### 3.2.6 densité en séquences

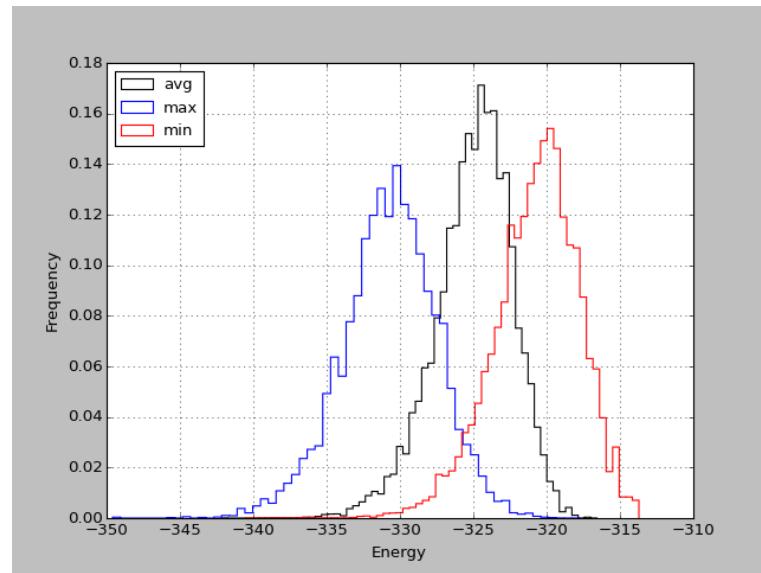


Figure 3.13 – .

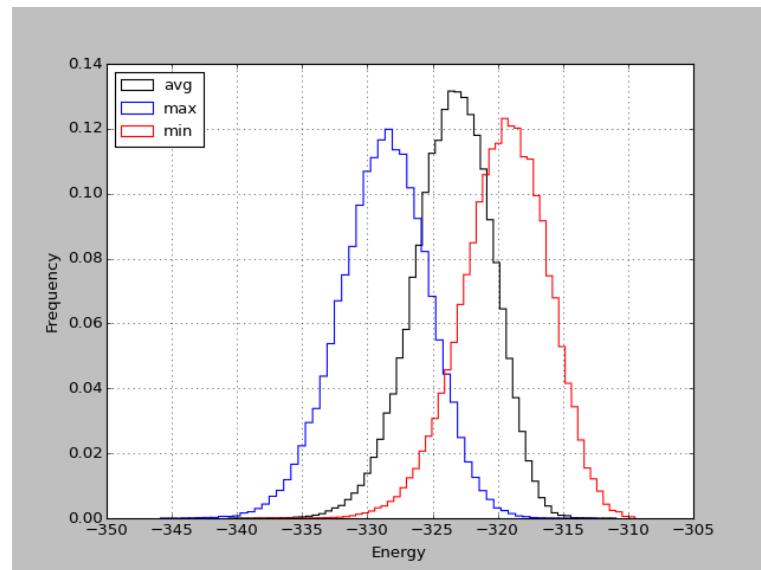


Figure 3.14 – .



# Conclusion

XXX



# **Titre d'une annexe**

XXX



# Bibliographie

Kofke D.A. (2002). Titre. *J. Chem. Phys.* **117**, 6911.

cité pages 3et 15

Kofke D.A. (2004a). Titre. *J. Chem. Phys.* **120**, 10852.

cité pages 3et 15

Kofke D.A. (2004b). Titre. *J. Chem. Phys.* **121**, 1167.

cité pages 3et 15



## Résum

**Titre de la thèse**

XXX

**Mots-clés :** motclé1, motclé2, motclé3

## Abstract

**Thesis title**

XXX

**Keywords:** keyword1, keyword2, keyword3