Alignement multiple

ClustalW

Romain GAUTIER

Polytech'Nice-Sophia

Département Génie Biologique

Que faire avec une séquence protéique ?

Rechercher des similitudes de séquence

Recherche d'autres séquences similaires dans les bases de données: BLAST sur la NR, Swiss-Prot…

Recherche de motifs fonctionnels: PROSITE, PFAM , BLOCKS,, PRINTS…

Prédire la structure

- Profil d'hydropatie, accessibilité au solvant
- Prédiction de structure secondaire
- Prédiction des hélices transmembranaires
- Prédiction des boucles (coudes)
- Modélisation par homologie
- Positionnement des chaînes latérales

Le contexte

Lorsque l'on compare une séquence contre une banque par BLAST, un grand nombre de séquences similaires peuvent être données en sortie. Cela nous permet de poser des hypothèses sur notre séquence inconnue (Fonction, …):

On peut ensuite regarder les alignements 2 à 2 entre les séquences similaires trouvées dans la banque et la séquence query:

Le contexte

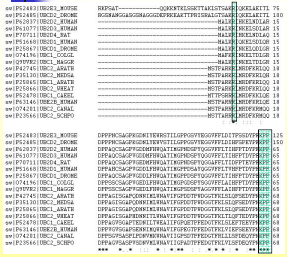
Mais, comment comparer toutes ces séquences et étudier leur relations ?

Comparer les séquences en une seule fois :

Alignement multiple

Un alignement multiple permet d'étudier un groupe de protéines ou de gènes apparentés, afin d'établir des relations d'évolution et mettre en évidence des motifs communs, jouant un rôle dans la fonction ou la structure de ces protéines homologues.

Exemple d'alignement multiple



Permet rapidement de retrouver les acides aminés strictement conservés dans toutes les séquences homologues, jouant souvent un rôle direct dans la fonction.

Un alignement multiple se conçoit à partir de 3 séquences et peut concerner jusqu'à plusieurs centaines de séquences.

Définition de l'alignement multiple

Entrée : k séquences (pas forcement la même longueur)

```
***************
**********
************
***********
```

Sortie: un tableau contenant les k séquences, avec des indels

```
***********_*****
***__**_**_**_**
***_*********__**
*****___**_*****
```

Démarche

Alignement 2 à 2

2 séquences quelconques

⇩

Détecter une similarité **syntaxique**

⇩

Il y a-t-il une fonction commune ?

Alignement multiple

Famille de séquences avec la même **fonction**

⇩

A quelle conservation syntaxique cela correspond-il ?

Exemple: Motif doigt de Zinc (JS Varré H Touzet)

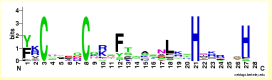
Voici un alignement multiple de séquences protéiques:

```
TYY1_HUMAN YVCPFDGCNKKFAQSTNLKSHIL--TH- 25
TF3B_BUFAM YRCPRENCDRTYTTKFNLKSHIL--TFH 26
ZN77_HUMAN YTCPEPHCGRGFTSATNYKNHVR--IH- 25
ZN76_HUMAN FRCGYKGCGRLYTTAHHLKVHER--AH- 25
P44_XENBO  YRCSYEDCQTVSPTWTALQTHLK--KH- 25
TSH_DROME  FRCVW--CKQSFPTLEALTTHMKDSKH- 25
XFIN_XENLA FRCSE--CSRSFTHNSDLTAHMR--KH- 23
EVI1_HUMAN YRCKY--CDRSFSISSNLQRHVRN-IH- 24
TRA1_CAEEL YKCEFADCEKAFSNASDRAKHQNR-TH- 26
TF3A_BUFAM CKCETENCNLAFTTASNMRLHFKR-AH- 26
SRYC_DROME FKCNY--CPRDFTNFPNWLKHTRR-RH- 24
ZG2-9_XENL FVCTV--CGKTYKYKHGLNTHLHS-H-- 23
ZG58_XENLA FVCTE--CNLSFAGLANLRSHQHL-H-- 23
YKQ8_CAEEL YKCTV--CRKDISSSESLRTHMFK-QHH 25
BASO_HUMAN FQCDI--CKKTFKNACSVKIHHKN-MH- 24
SUHW_DROAN YACKI--CGKDFTRSYHLKRHQKY-SSC 25
ZN10_HUMAN YKCNQ--CGIIFSQNSPFIVHQI--AH- 23
P43_XENBO  LKCSVPGCKRSFRKKRALRIHVS--EH- 25
IKAR_MOUSE FECNM--CGYHSQDRYEFSSHITRGEH- 25
```

Exemple: Motif doigt de Zinc (JS Varré H Touzet)

Regardons les résidus totalement conservés:

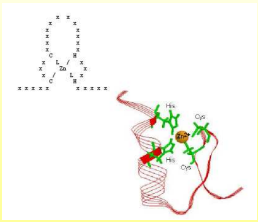
```
TYY1_HUMAN YVCPFDGCNKKFAQSTNLKSHIL--TH- 25
TF3B_BUFAM YRCPRENCDRTYTTKFNLKSHIL--TFH 26
ZN77_HUMAN YTCPEPHCGRGFTSATNYKNHVR--IH- 25
ZN76_HUMAN FRCGYKGCGRLYTTAHHLKVHER--AH- 25
P44_XENBO  YRCSYEDCQTVSPTWTALQTHLK--KH- 25
TSH_DROME  FRCVW--CKQSFPTLEALTTHMKDSKH- 25
XFIN_XENLA FRCSE--CSRSFTHNSDLTAHMR--KH- 23
EVI1_HUMAN YRCKY--CDRSFSISSNLQRHVRN-IH- 24
TRA1_CAEEL YKCEFADCEKAFSNASDRAKHQNR-TH- 26
TF3A_BUFAM CKCETENCNLAFTTASNMRLHFKR-AH- 26
SRYC_DROME FKCNY--CPRDFTNFPNWLKHTRR-RH- 24
ZG2-9_XENL FVCTV--CGKTYKYKHGLNTHLHS-H-- 23
ZG58_XENLA FVCTE--CNLSFAGLANLRSHQHL-H-- 23
YKQ8_CAEEL YKCTV--CRKDISSSESLRTHMFK-QHH 25
BASO_HUMAN FQCDI--CKKTFKNACSVKIHHKN-MH- 24
SUHW_DROAN YACKI--CGKDFTRSYHLKRHQKY-SSC 25
ZN10_HUMAN YKCNQ--CGIIFSQNSPFIVHQI--AH- 23
P43_XENBO  LKCSVPGCKRSFRKKRALRIHVS--EH- 25
IKAR_MOUSE FECNM--CGYHSQDRYEFSSHITRGEH- 25
```



On peut établir une carte par position des résidus conservées :
WebLogo : http://weblogo.berkeley.edu/logo.cgi

Exemple: Motif doigt de Zinc (JS Varré H Touzet)

```
TYY1_HUMAN YVCPFDGCNKKFAQSTNLKSHIL--TH- 25
TF3B_BUFAM YRCPRENCDRTYTTKFNLKSHIL--TFH 26
ZN77_HUMAN YTCPEPHCGRGFTSATNYKNHVR--IH- 25
ZN76_HUMAN FRCGYKGCGRLYTTAHHLKVHER--AH- 25
P44_XENBO  YRCSYEDCQTVSPTWTALQTHLK--KH- 25
TSH_DROME  FRCVW--CKQSFPTLEALTTHMKDSKH- 25
XFIN_XENLA FRCSE--CSRSFTHNSDLTAHMR--KH- 23
EVI1_HUMAN YRCKY--CDRSFSISSNLQRHVRN-IH- 24
TRA1_CAEEL YKCEFADCEKAFSNASDRAKHQNR-TH- 26
TF3A_BUFAM CKCETENCNLAFTTASNMRLHFKR-AH- 26
SRYC_DROME FKCNY--CPRDFTNFPNWLKHTRR-RH- 24
ZG2-9_XENL FVCTV--CGKTYKYKHGLNTHLHS-H-- 23
ZG58_XENLA FVCTE--CNLSFAGLANLRSHQHL-H-- 23
YKQ8_CAEEL YKCTV--CRKDISSSESLRTHMFK-QHH 25
BASO_HUMAN FQCDI--CKKTFKNACSVKIHHKN-MH- 24
SUHW_DROAN YACKI--CGKDFTRSYHLKRHQKY-SSC 25
ZN10_HUMAN YKCNQ--CGIIFSQNSPFIVHQI--AH- 23
P43_XENBO  LKCSVPGCKRSFRKKRALRIHVS--EH- 25
IKAR_MOUSE FECNM--CGYHSQDRYEFSSHITRGEH- 25
```



L'alignement multiple permet de mieux comprendre l'organisation d'un motif particulier et permet de modéliser le motif : (Prosite)
C-x(2,4)-C-x(3)-[LIVMFYWC]-x(8)-H-x(3,5)-H

Notion de séquence consensus

A partir de l'alignement multiple, on peut déterminer la séquence consensus:

On attribut à chaque position l'acide aminé ou le nucléotide qui est le plus souvent retrouvé (totalement conservé ou partiellement).

```
                    130       140       150       160       170       180
                     |         |         |         |         |         |
UB2E3_MOUSE TILGPPGSVYEGGVFFLDITFSSDYPFKPPKVTFRTRIYHCNISQGVICLDILKDNWSP
UBCD2_DROME TILGPPGSVYEGGVFFLDIHFSPEYPFKPPKVTFRTRIYHCNISQGVICLDILKDNWSP
UB2D2_HUMAN TIMGPNDSPYQGGVFFLTIHFPTDYPFKPPKVAFTTRIYHPNINSNGSICLDILRSQWSP
UB2D3_HUMAN TIMGPNDSPYQGGVFFLTIHFPTDYPFKPPKVAFTTRIYHPNINSNGSICLDILRSQWSP
UB2D4_RAT   TIMGPNDSPYQGGAFFLTIDFPTEYPFKPPKVEFTTRIYHPNVNSNGSICLDILRSQWSP
UB2D1_HUMAN TIMGPPDSAYQGGVFFLTVHFPTDYPFKPPKIAFTTKIYHPNINSNGSICLDILRSQWSP
UBCD1_DROME TIMGPPDSPYQGGVFFLTIHFPTDYPFKPPKVAFTTRIYHPNINSNGSICLDILRSQWSP
UBC1_COLGL  TIMGPGDSPYSGGVFFLAIHFPTDYPFKPPKVNFTTRIYHPNINSNGSICLDILRDQWSP
UBC1_MAGGR  TIMGPSDSPYAGGVFFLAIHFPTDYPFKPPKVNFTTRIYHPNINSNGSICLDILRDQWSP
UBC2_ARATH  VIFGPDDTPWDGGTFKLSLQFSEDYPNKPPTVRFVSRMFHPNIYADGSICLDILQNQWSP
UBC2_MEDSA  VIFGPDDTPWDGGTFKLSLQFSEDYPNKPPTVRFVSRMFHPNIYADGSICLDILQNQWSP
UBC1_ARATH  VIFGPDDTPWDGGTFKLSLQFSEDYPNKPPTVRFVSRMFHPNIYADGSICLDILQNQWSP
UBC2_WHEAT  VIFGPDDTPWDGGTFKLTLQFTEDYPNKPPTVRFVSRMFHPNIYADGSICLDILQNQWSP
UBC1_CAEEL  IIFGPQETPFEDGTFKLSLEFTEEYPNKPPTVKFISKMFHPNVYADGSICLDILQNRWSP
UBE2B_HUMAN VIFGPEGTPFEDGTFKLVIEFSEEYPNKPPTVRFLSKMFHPNVYADGSICLDILQNRWSP
UBC2_CANAL  VIIGPSDTPFEDGTFRLLLSFDEQYPNKPPQVKFISEMFHPNVYASGELCLDILQNRWSP
UBC2_SCHPO  VIIGPADTPFEDGTFKLVLSFDEQYPNKPPLVKFVSTMFHPNVYANGELCLDILQNRWSP
                 .:***  : . .* * . * .  .** *** . .  . : ::* * : .* :*****.. ***
Prim.cons.  TIMGP2DSPYEGG2FFLTIHFPEDYPFKPPKVRFTTRIYHPNINSNGSICLDILQNQWSP
```

Alignement multiple: principe

L'approche la plus courante consiste à aligner progressivement des paires de séquences:

Alignements itératifs et progressifs

Algorithme Needleman & Wunsh 2 à 2 traditionnel répété n fois

Alignement multiple: principe

L'approche la plus courante consiste à aligner progressivement des paires de séquences.

Les différentes approches se distinguent par:
La façon de choisir la paire initiale de séquences

Pour progresser dans l'alignement, les programmes vont:
Soit aligner chaque séquence les unes après les autres à un alignement unique enrichi à chaque étape
Soit créer des sous-familles de séquences d'abord alignées au sein de ces familles puis entre les familles.

La méthode de pondération des alignements individuels des paires de séquences et des alignements cumulés.

Alignement basé sur un arbre

Idée: reconstruire l'alignement multiple à partir d'un **arbre guide** (clusters)
feuilles : séquences
noeuds: alignements

Partir des feuilles puis remonter dans l'arbre
Utilisation de la technique de **profil alignement** -> produire un seul alignement multiple avec deux.

CLUSTer + **AL**ignement **CLUSTAL**

Thompson et al. 1994

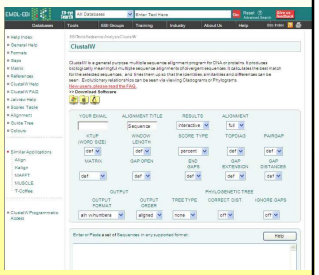
ClustalW EBI http://www.ebi.ac.uk/clustalw/index.html

ClustalW est l'un des programmes les plus utilisés pour l'alignement progressif.

Etape 1: Alignements globaux 2 à 2

Etape 2: Regroupements des alignements (clusters), construction arbre guide

Etape 3: Alignement multiple obtenu par combinaisons des alignements 2 à 2 (profils)



ClustalW exemple

4 séquences

s1 cgatgagtcattgtgactg
s2 cgagccattgtagctactg
s3 cgaccattgtagctacctg
s4 cgatgagtcactgtgactg

Jeu de score:
Indel= -2
Substitution=-1
Identité= 1

ClustalW étape 1: Calcul des scores

Les alignements de toutes les paires de séquences sont réalisés puis le programme génère une matrice de distances décrivant leur taux de similitude.

```
s1      cgatgagtcattgt-g--actg          s2      cgagccattgtagcta-ctg
        ||| |   |||||| |  ||||                  ||| ||||||||||||| |||
s2      cga-g--ccattgtagctactg          s3      cga-ccattgtagctacctg

s1      cgatgagtcattg-tgactg            s2      cga-g--ccattgtagctactg
        |||  | |  | | |  |||                     ||| |   || ||| |  ||||
s3      cgacca-ttgtagctacctg            s4      cgatgagtcactgt-g--actg

s1      cgatgagtcattgtgactg             s3      cgaccattgtagctacctg
        ||||||||||| ||||||||                    |||  | |     |  |||
s4      cgatgagtcactgtgactg             s4      cgatgagtcactgtgactg
```

Tableau des scores d'alignement:

	s1	s2	s3	S4
s1		2	0	17
s2	2		14	0
s3	0	14		-1
s4	17	0	-1	

Pour N séquences:

N(N-1)/ 2 calculs

ClustalW étape 2: Construction arbre

A l'aide de la matrice de scores, un arbre guide (ou dendogramme) est ensuite construit par un algorithme dit de neighbor-joining:

	s1	s2	s3	S4
s1		2	0	17
s2	2		14	0
s3	0	14		-1
s4	17	0	-1	

s1
s2
s3
s4

Regroupement des séquences suivant leur similitude à partir de la matrice des scores 2 à 2.

ClustalW étape 2: Construction arbre

A l'aide de la matrice de scores, un arbre guide (ou dendogramme) est ensuite construit par un algorithme dit de neighbor-joining:

	s1	s2	s3	S4
s1		2	0	17
s2	2		14	0
s3	0	14		-1
s4	17	0	-1	

s1

s2

s3

s4

ClustalW étape 2: Construction arbre

A l'aide de la matrice de scores, un arbre guide (ou dendogramme) est ensuite construit par un algorithme dit de neighbor-joining:

	s1	s2	s3	S4
s1		2	0	17
s2	2		14	0
s3	0	14		-1
s4	17	0	-1	

s1

s2

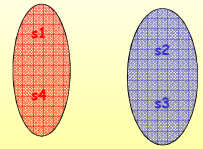
s3

s4

ClustalW étape 2: Construction arbre

A l'aide de la matrice de scores, un arbre guide (ou dendogramme) est ensuite construit par un algorithme dit de neighbor-joining:

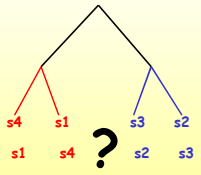
	s1	s2	s3	S4
s1		2	0	17
s2	2		14	0
s3	0	14		-1
s4	17	0	-1	



ClustalW étape 2: Construction arbre

A l'aide de la matrice de scores, un arbre guide (ou dendogramme) est ensuite construit par un algorithme dit de neighbor-joining:

	s1	s2	s3	S4
s1		2	0	17
s2	2		14	0
s3	0	14		-1
s4	17	0	-1	



ClustalW étape 2: Construction arbre

A l'aide de la matrice de scores, un arbre guide (ou dendogramme) est ensuite construit par un algorithme dit de neighbor-joining:

	s1	s2	s3	S4
s1		2	0	17
s2	2		14	0
s3	0	14		-1
s4	17	0	-1	

Alignement s1s2 est plus proche que s1s3

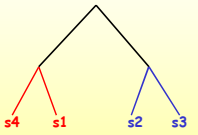
Alignement s4s2 est plus proche que s4s3

Matrice de score -> matrice de distance

ClustalW étape 2: Construction arbre

A l'aide de la matrice de scores, un arbre guide (ou dendogramme) est ensuite construit par un algorithme dit de neighbor-joining:

	s1	s2	s3	S4
s1		2	0	17
s2	2		14	0
s3	0	14		-1
s4	17	0	-1	



Arbre guide construit

Etape 3: Construction alignement final

ClustalW aligne les séquences en se servant de l'arbre guide: chaque paire de séquences situées sur une même branche extérieure de l'arbre est alignée par programmation dynamique.

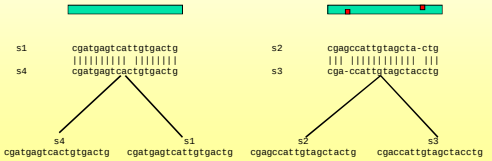
Les alignements partiels permettent de constituer des **profils**, représentés sous forme de tableau dans lequel sont données pour chaque position la fréquence observée de chaque lettre.

L'algorithme aligne ensuite les profils associés par un même nœud de l'arbre. Cet alignement de séquences puis de profils se poursuit de façon récursive jusqu'à l'alignement final complet depuis les branches de l'arbre vers la racine.

Etape 3: Construction alignement final

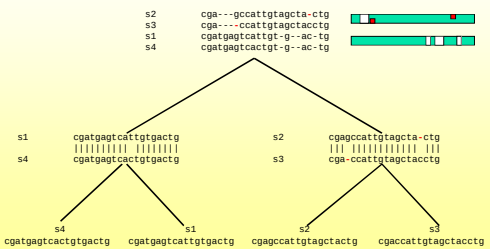
L'alignement et création des profils:

Etape 3: Construction alignement final



Etape 3: Construction alignement final

Once a gap, always a gap



ClustalW

ClustalW est optimisé pour les protéines:

Pondération des séquences en fonction de leur sur/sous représentation

Adaptation des matrices de similitudes au fil de l'algorithme en fonction de la divergence des séquences à aligner
- Blosum 80 pour aligner des séquences proches
- Blosum 50 pour aligner des séquences distantes

Pénalités de gaps spécifiques à chaque résidu.
Par exemple, les Glycines sont davantage susceptible d'avoisiner un gap que les Valines.

Pénalités de gaps réduites dans les régions hydrophiles
Encourage la formation de gaps dans des boucles plutôt que dans des régions structurées.

Pénalités de gaps augmentées dans le voisinage d'autres gaps
Evite la formation de petits gaps voisins, au profit de longs gaps.

Autre méthode

A partir des alignements locaux

Idée: repérer des similitudes locales fortes entre les séquences (les diagonales du dotplot par exemple)

Incorporer les diagonales dans l'alignement multiple

Conséquence: les gaps inter-diagonales sont moins importants

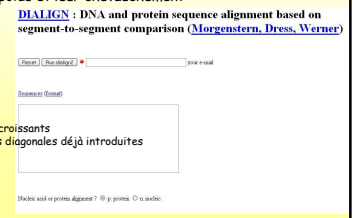
DIagonal + **ALIGN**ement **DIALIGN**

DIALIGN Morgenstern et al. 1996

DIALIGN sur Pasteur: http://bioweb.pasteur.fr/seqanal/interfaces/dialign2-simple.html

Alignements des paires avec optimisation des poids des diagonales

Tri des diagonales selon leur poids et leur chevauchement



DIALIGN : **DNA and protein sequence alignment based on segment-to-segment comparison (Morgenstern, Dress, Werner)**

Nucleic acid or protein alignment ? ● p: protein ○ n: nucleic

Reconstruction gloutonne

Insertion des diagonales par poids décroissants

Vérification de la consistance avec les diagonales déjà introduites

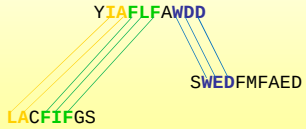
Recommencer …

DIALIGN exemple

3 séquences:

YIAFLFAWDD
LACFIFGS
SWEDFMFAED

Etape 1 : Détection des diagonales dans les paires de séquences:

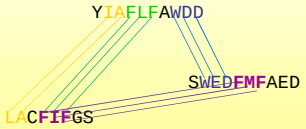


DIALIGN exemple

3 séquences:

YIAFLFAWDD
LACFIFGS
SWEDFMFAED

Etape 1 : Détection des diagonales dans les paires de séquences:

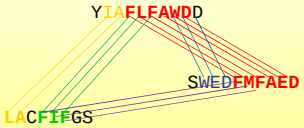


DIALIGN exemple

3 séquences:

YIAFLFAWDD
LACFIFGS
SWEDFMFAED

Etape 1 : Détection des diagonales dans les paires de séquences:



DIALIGN exemple

3 séquences:

YIAFLFAWDD
LACFIFGS
SWEDFMFAED

Etape 2 : Sélection d'un ensemble cohérent de diagonales pour construire l'alignement:

Pas de croisement
Pas de chevauchements
score maximal

```
yIA-FLFAWDd
-LAcFIFgs--
swedFMFAED-
```

CLUSTAL vs DIALIGN

Exemple (C. Notre-Dame)

GARFIELD THE LAST FAT CAT

GARFIELD THE FAT CAT

GARFIELD THE VERY FAST CAT

THE FAT CAT

CLUSTAL vs DIALIGN

Alignement fourni par ClustalW:

```
seq2xx1 GARFIELDTHE----FAT-CAT
seq4xx3 --------THE----FAT-CAT
seq1xx0 GARFIELDTHELASTFAT-CAT
seq3xx2 GARFIELDTHEVERYFASTCAT
```

Clustal 1ère version:

```
GARFIELDTHELASTFA-TCAT
----GARFIELDTHEFA-TCAT
GARFIELDTHEVERYFASTCAT
-----------THEFA-TCAT
```

Alignement fourni par Dialign2:

```
seq1 1 GARFIELDTHELASTFA-TCAT
seq2 1 GARFIELDTHE----FA-TCAT
seq3 1 GARFIELDTHEVERYFASTCAT
seq4 1 --------THE----FA-TCAT
```

Quelles méthode utiliser

Cela dépend du type de séquences à aligner !

BaliBase : base de données d'alignements multiples et de benchmarks
http://www-bio3d-igbmc.u-strasbg.fr/balibase/

Plusieurs références!
Ref1 : Séquences équidistantes avec différents niveaux de conservation
Ref2: Protéines homologues + une séquence orpheline
Ref3: Sous-groupes avec moins de 25% d'identité entre les groupes
Ref4: Extensions N/C terminales
Ref5: Insertions internes
Ref6: Répétitions internes
Ref7: Protéines transmembranaires
Ref8: Permutations de domaines

Ref1, 2 3: Préférer Clustal à Dialign
Ref4 et 5: Préférer Dialign à Clustal

Quelles méthode utiliser

Plus les séquences sont divergentes, moins le résultat est fiable.
Quand le taux d'identité est supérieur à 35%, toutes les méthodes sont satisfaisantes.

Twilight Zone : 10-20% d'identité
Aucune méthode n'assure un alignement avec plus de 50% de correction.

Clustal a tendance à autoriser moins de gaps que Dialign

Similitude locale : Dialign
Similitude globale : Clustal

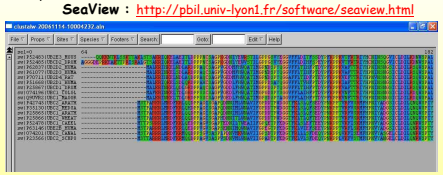
Existe d'autres méthodes : MultiAlign, Tcoffee, etc…

Pas de méthode universelle
Pas de confiance aveugle vis-à-vis du résultat obtenu

Visualiser des alignements multiples

Il existe des outils simples pour visualiser ou modifier un alignement multiple:

SeaView : http://pbil.univ-lyon1.fr/software/seaview.html



BioEdit : http://www.mbio.ncsu.edu/BioEdit/bioedit.html

Autre type d'alignement: structuraux

On peut aussi vouloir aligner non pas des séquences mais des structures protéiques (Cela permet de mieux comprendre les zones clés, structurellement conservées).

TM-Align : http://zhang.bioinformatics.ku.edu/TM-align/
Permet d'aligner 2 structures au format PDB

Isuperpose : http://bioserv.rpbs.jussieu.fr/cgi-bin/iSuperpose
Permet d'aligner plusieurs structures (format PDB). Possibilité de fournir l'alignement ou pas.

../..

Exercises

Exo 1: A partir du jeu de séquences seq1, lancer un ClustalW et analyser l'alignement obtenu.

Regarder les différentes sorties proposées par ClustalW (lancer l'applet JalView)

Sauvegarder les différents fichiers de sortie.
Télécharger et installer SeaView et Njplot.

Regarder l'alignement avec SeaView et visualiser l'arbre guide avec NjPlot.

Exercises

Exo 2: Analyse des domaines SH3
Intro: SH3 (Src homology 3) domains are often indicative of a protein involved in signal transduction related to cytoskeletal organization. The SH3 domain has a characterisitic fold which consists of five or six beta-strands arranged as two tightly packed anti-parallel beta sheets.
The linker regions may contain short helices.

Voici 5 protéines à aligner: seq2
Ce sont des séquences courtes avec des similitudes faibles et diffuses (<25%).

Tester les différents alignements obtenus avec ClustalW, Dialign2, Multialign.
Proposer des régions ou résidus clés pour ce type de domaine.

Exercises

Exo 2: Les structures de ces protéines sont connues:
Voici pour chaque protéine les éléments de structures secondaires:

```
> 1aboA 58
NLFVALYDFVASGDNTLSITKGEKLRVLGYNHNGEWCEAQTKNGQGWVPSNYITPVNS
CEEEECCCCCCCCCCCCCCCCEEEEEEEECCCCEEEEEECCEEEEEEHHHEEECCC
> 1ycsB 60
KGVIYALWDYEPQNDDELPMKEGDCMTIIHREDEDEIEWWWARLNDKEGYVPRNLLGLYP
CCEEEECCCCCCCCCCCCCCCCEEEEEECCCCCCCCCEEEEEECCEEEEEEHHHEECCC
> 1pht 83
AEGYQYRALYDYKKEREEDIDLHLGDILTVNKGSLVALGFSDGQEARPEEIGWLNGYNETTGERGDFPGTYVEYIGRKKISPP
CEEEEEEECCCCCCCCCCCCCCCCCCEEEEEEHHHHHHCCCCCCHHHCHHHHCEEEEEECCCCEEEEEECCCEEEEEEECCCC
> 1ihvA 52
MIQNFRVYYRDSRDPVWKGPAKLLWKGEGAVVIQDNSDIKVVPRRKAKIIRD
CCCCEEEEEEECCEEEEEEEEEEEECCCEEEEECCCCEEEEEHHHEEEEEC
> 1vie 60
PSNATFGMGDRVRKKSGAAWQGQIVGWYCTNLTPEGYAVESEAHPGSVQIYPVAALERIN
CCCCCCCCCEEEEECCCCCCEEEEEEEEECCEEEEEEEECCCCCCEEEEEHHEEECC
```

Voici les protéines alignées structurellement visualisées avec Pymol fit

Que peut-on dire au niveau de la structure ?
(par rapport à l'alignement ?)