

Alignement multiple

Hélène TOUZET

touzet@lifl.fr

Alignement multiple

▷ **Entrée** : k séquences

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

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

```
* * * * * * * * * - * * * *
* * * - - - * * * - * * * *
* * * - * * * * * * * * *
* * * - - * * - - * * * * *
```

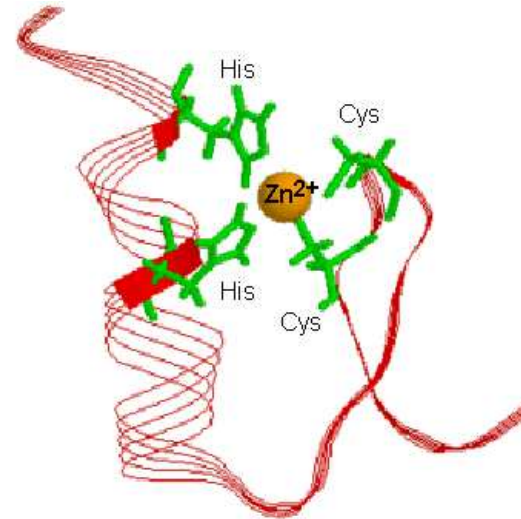
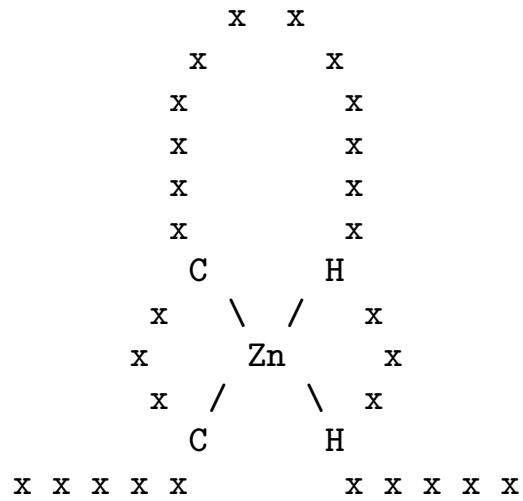
Exemple : Doigt de zinc

YVCPFDGCN---KKFAQSTNLKSHILT---H
YKCT--VCR---KDISSESRLRTHMFKQ-HH
FQCD--ICK---KTFKNACSVKIHHKN--MH
FVCT--VCG---KTYKYKHGLNTHLHS---H
LKCSVPGCK---RSFRKKRALRIHVSE---H
FECN--MCG---YHSQDRYEFSSHITRG-EH
YTCG--YCTEDSPSPRPSLLESHISL--MH
YKCEFADCE---KAFSNASDRAKHQNR--TH
YKCN--QCG---IIFSQNSPFIVHQIA---H
FVCHWQDCSRELRPFKAQYMLVVHMRR---H
FRCS--ECS---RSFTHNSDLTAHMRK---H
CKCETENCN---LAFTTASNMLRHFKR--AH
FVCT--ECN---LSFAGLANLRSHQHL---H
YRCSYEDCQ---TVSPTWTALQTHLKK---H
FRCV--WCK---QSFPLEALTTHMKDS-KH
FRCGYKCG---RLYTTAHLKVHERA---H
YRCPRENCN---RTYTTKFNKSHILT--FH
YACK--ICG---KDFTRSYHLKRHQKYS-SC
YTCPEPHCG---RGFTSATNYKNHVRI---H

Exemple : Doigt de zinc

YVCPFDGCN---KKFAQSTNLKSHILT---H
YKCT--VCR---KDISSESRLRTHMFKQ-HH
FQCD--ICK---KTFKNACSVKIHHKN--MH
FVCT--VCG---KTYKYKHGLNTHLHS---H
LKCSVPGCK---RSFRKKRALRIHVSE---H
FECN--MCG---YHSQDRYEFSSHITRG-EH
YTCG--YCTEDSPSPRPSLLESHISL--MH
YKCEFADCE---KAFSNASDRAKHQNR--TH
YKCN--QCG---IIFSQNSPFIVHQIA---H
FVCHWQDCSRELRPFKAQYMLVVMR---H
FRCS--ECS---RSFTHNSDLTAHMRK---H
CKCETENCN---LAFTTASNMRLEHFKR--AH
FVCT--ECN---LSFAGLANLRSHQHL---H
YRCSYEDCQ---TVSPTWTALQTHLKK---H
FRCV--WCK---QSFPTLEALTTHMKDS-KH
FRCGYKCG---RLYTTAHLKVMHERA---H
YRCPRENCN---RTYTTKFNLSHILT--FH
YACK--ICG---KDFTRSYHLKRHQKYS-SC
YTCPEPHCG---RGFTSATNYKNHVRI---H

C-x(2,4)-C-x(3)-[LIVMFYWC]-x(8)-H-x(3,5)-H



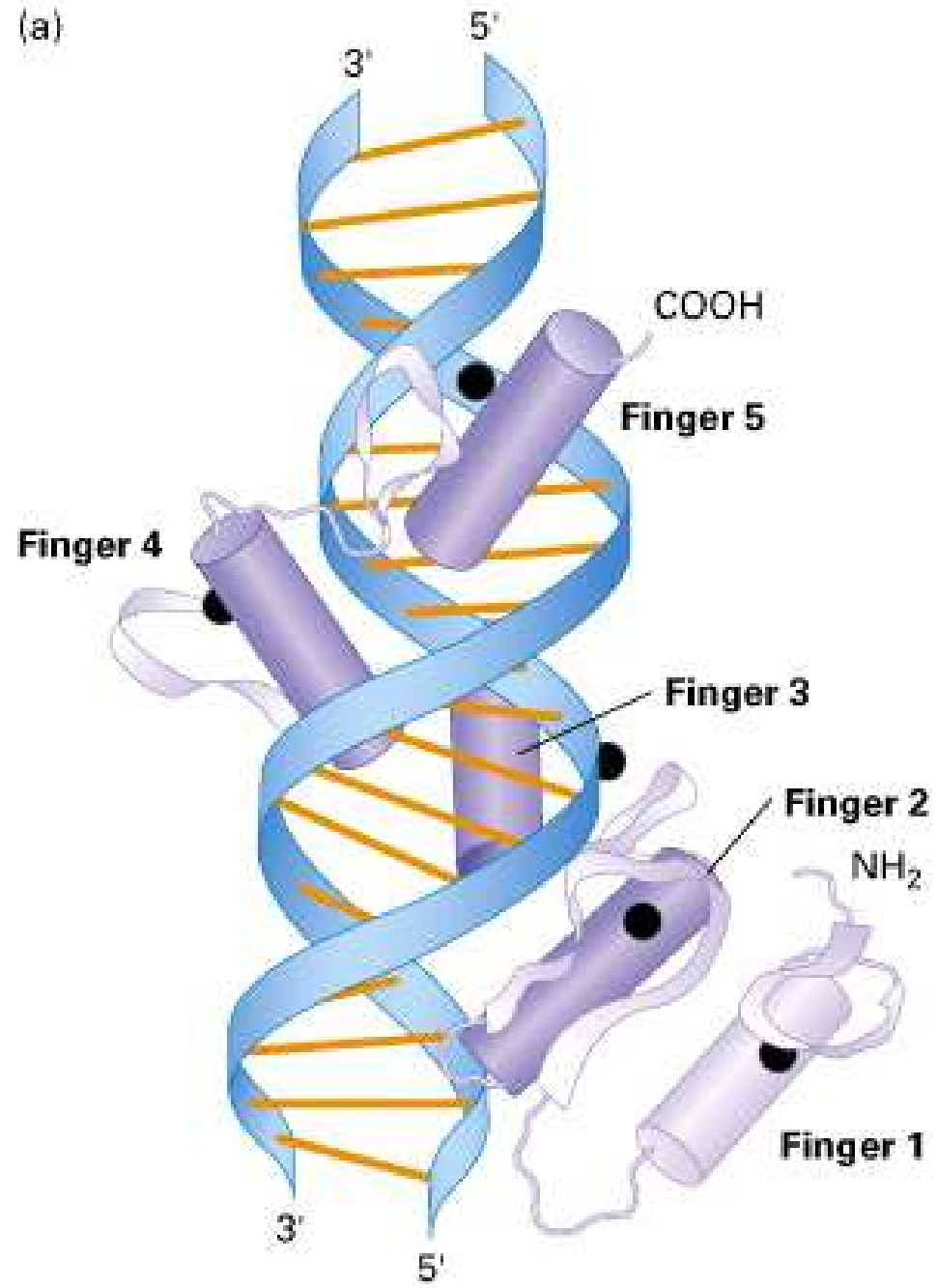
Modélisation

Prosite : motif

C-x(2,4)-C-x(3)-[LIVMFYWC]-x(8)-H-x(3,5)-H

Pfam : modèle de Markov caché (HMM)

(a)



Exemple : Site de fixation de la cellulose

HWGQCGGI---GYSGCKTCTSGTTCQYSNDYYSQCL
HYGQCGGI---GYSGPTVCASGTTCQVLNPYYSQCL
QWGQCGGI---GYTGSTTCASPYTCHVLNPYYSQCY
VWGQCGGQ---NWSGPTCCASGSTCVYSNDYYSQCL
LYGQCGGA---GWTGPTTCQAPGTCKVQNWYSQCL
IWGQCGGN---GWTGATTCASGLKCEKINDWYYQCV
VWGQCGGN---GWTGPTTCASGSTCVKQNDFYSQCL
DWAQCGGN---GWTGPTTCVSPYTCTKQNDWYSQCL
QWGQCGGQ---NYSGPTTCKSPFTCKKINDFYSQCQ
RWQQCGGI---GFTGPTQCEEPYICTKLNDWYSQCL
HWAQCGGI---GFSGPTTCPEPYTCAKDHDYYSQCV
LYEQCGGI---GFDGVTCCSEGLMCMKMGYPYYSQCR
VWAQCGGQ---NWSGTPCCTSGNKCCKLNDYYSQCC
PYGQCGGM---NYSGKTMCSFGKCVLNEFFSQCD
AYYQCGGSKSAYPNGNLACATGSKCVKQNEYYSQCV
EYAACGGE---MFMGAKCCKFGLVCYETSGKWSQCR

Exemple : Site de fixation de la cellulose

HWGQCGGI---GYSCKTCTSGTTCQYSNDYYSQCL
HYGQCGGI---GYSPTVCASGTTQVLNPYYSQCL
QWGQCGGI---GYTGSTTCASPYTCHVLNPYYSQCY
VWGQCGGQ---NWSGPTCCASGSTCVYSNDYYSQCL
LYGQCGGA---GWTGPTTCQAPGTCKVQNPYYSQCL
IWGQCGGN---GWTGATTCASGLKCEKINDWYYQCV
VWGQCGGN---GWTGPTTCASGSTCVKQNDYYSQCL
DWAQCGGN---GWTGPTTCVSPYTCTKQNDWYSQCL
QWGQCGGQ---NYSGPTTCCKSPFTCKKINDYYSQCQ
RWQQCGGI---GFTGPTQCEEPYICTKLNDWYSQCL
HWAQCGGI---GFSGPTTCPEPYTCAKDHDYYSQCV
LYEQCGGI---GFDGVTCCSEGLMCMKMGPYYSQCR
VWAQCGGQ---NWSGTPCCTSGNKCVKLNDYYSQCQ
PYGQCGGM---NYSGKTM CSPGFKCVELNEFFSQCD
AYYQCGGSKSAYPNGNLACATGSKCVKQNEYYSQCV
EYAA CGGE---MFMGAKCCKFGLVCYETSGKWSQCR

C-G-G-x(4,7)-G-x(3)-C-x(5)-C-x(3,5)-[NHG]-x-[FYWM]-x(2)-Q-C

Les 4 cystéines sont impliquées dans des liaisons di-sulfures.

Alignement 2 à 2

2 séquences quelconques



détection d'une similarité

syntaxique



? **fonction** commune ?

Alignement multiple

Famille de séquences

avec la même **fonction**



A quelle conservation

syntaxique

cela correspond-il ?

Comment scorer un alignement multiple

Le score SP (Sum of Pairs)

- ▷ Score de l'alignement : somme des scores de ses colonnes
- ▷ Comment scorer une colonne ?
 - adaptable à un nombre quelconque d'arguments
 - indépendant de l'ordre
 - reflète la similarité

$$\text{scoreSP} \begin{pmatrix} s_1 \\ \vdots \\ s_k \end{pmatrix} = \sum_{1 \leq i < j \leq k} \text{score}(s_i, s_j)$$

$s_1, \dots, s_k \in \mathcal{A} \cup \{-\}$ et $\text{score}(-, -) = 0$

A	A	C	G	T	A	C	G	A	T	A
A	-	C	G	T	A	-	A	A	T	G
G	T	C	G	T	A	-	-	T	T	A

Identité : +1,
Substitution: -2,
Indel : -3

Définition alternative (équivalente)

- ▷ α : alignement multiple pour les séquences s_1, \dots, s_k
- ▷ α_{ij} : projection de l'alignement pour s_i et s_j

$$\text{scoreSP}(\alpha) = \sum_{1 \leq i < j \leq k} \text{score}(\alpha_{ij})$$

- ▷ Retour à l'exemple

A	A	C	G	T	A	C	G	A	T	A
A	-	C	G	T	A	-	A	A	T	G
G	T	C	G	T	A	-	-	T	T	A

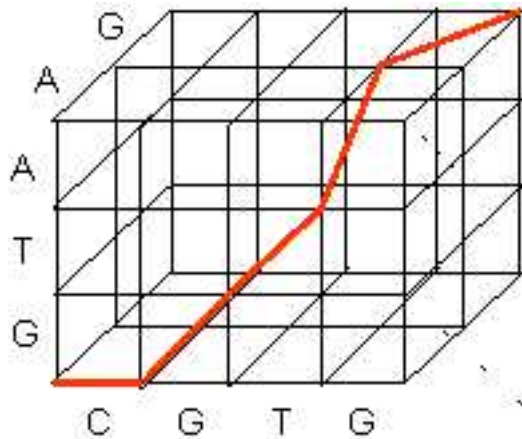
Identité : +1,
Substitution: -2,
Indel : -3

Problème algorithmique

Trouver l'alignement multiple de score SP maximal.

Approche exacte : programmation dynamique

- ▷ Alignement deux à deux : chemin dans une matrice de dimension 2
- ▷ Alignement multiple : chemin dans une matrice de dimension supérieure



C	G	T	-	G
-	G	T	A	-
-	-	-	A	G

Complexité

s_1, \dots, s_k : séquences de taille n

$T(i_1, \dots, i_k)$: score optimal entre les k
préfixes $s_1(1..i_1), \dots, s_k(1..i_k)$

- ▷ Table de taille n^k
- ▷ Temps de calcul d'une case : dépend de $2^k - 1$ cases précédentes
- ▷ Temps de calcul de chaque scoreSP candidat : $k(k - 1)/2$

$$O(n^k 2^k k^2)$$

Complexité

s_1, \dots, s_k : séquences de taille n

$T(i_1, \dots, i_k)$: score optimal entre les k
préfixes $s_1(1..i_1), \dots, s_k(1..i_k)$

- ▷ Table de taille n^k
- ▷ Temps de calcul d'une case : dépend de $2^k - 1$ cases précédentes
- ▷ Temps de calcul de chaque scoreSP candidat : $k(k - 1)/2$

$$O(n^k 2^k k^2)$$

Le problème de décision associé est NP-complet

Recours à des programmes heuristiques :

- ▷ Clustal (le plus populaire)
- ▷ Dialign2 (complémentaire à Clustal)
- ▷ T-coffee, Pima, Multalin, . . .

Autant d'alignements que de programmes

Le programme Clustal

Thompson *et al.* - 1994

CLUSTAL = cluster + alignement

Étape 1 : alignements globaux 2 à 2

Étape 2 : regroupement des alignements (clusters)

Étape 3 : alignement multiple obtenu par combinaisons des alignements 2 à 2

Exemple :

```
s1    cgatgagtcattgtgactg
s2    cgagccattgtagctactg
s3    cgaccattgtagctacctg
s4    cgatgagtcactgtgactg
```

indel : -2, substitution : -1, identité : 1

Étape 1 : *calcul des scores de similarité de tous les alignements globaux 2 à 2*

s1 cgatgagtcattgt-g--actg
 ||| | ||||| | ||||
 s2 cga-g--ccattgtagctactg

s2 cgagccattgtagcta-ctg
 ||| ||||| ||||| |||
 s3 cga-ccattgtagctacctg

s1 cgatgagtcattg-tgactg
 ||| | | | | |||
 s3 cgacca-ttgtagctacctg

s2 cga-g--ccattgtagctactg
 ||| | || ||| | ||||
 s4 cgatgagtcactgt-g--actg

s1 cgatgagtcattgtgactg
 ||||| ||||| |||||
 s4 cgatgagtcactgtgactg

s3 cgaccattgtagctacctg
 ||| | | | |||
 s4 cgatgagtcactgtgactg

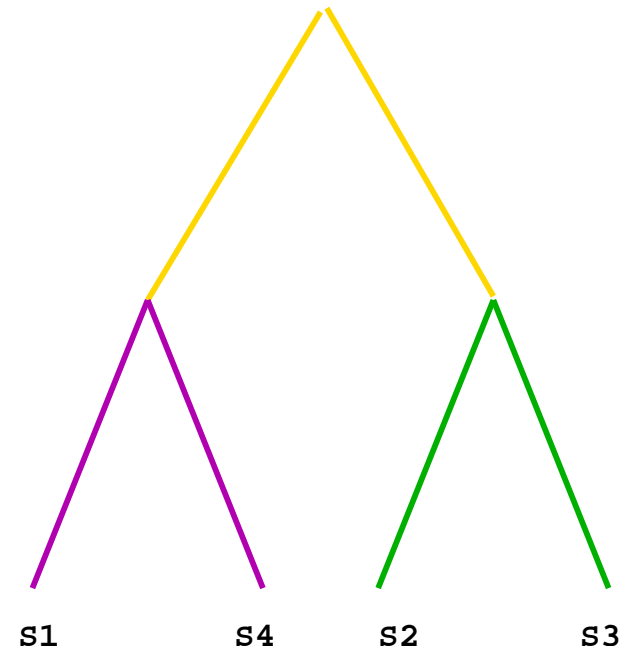
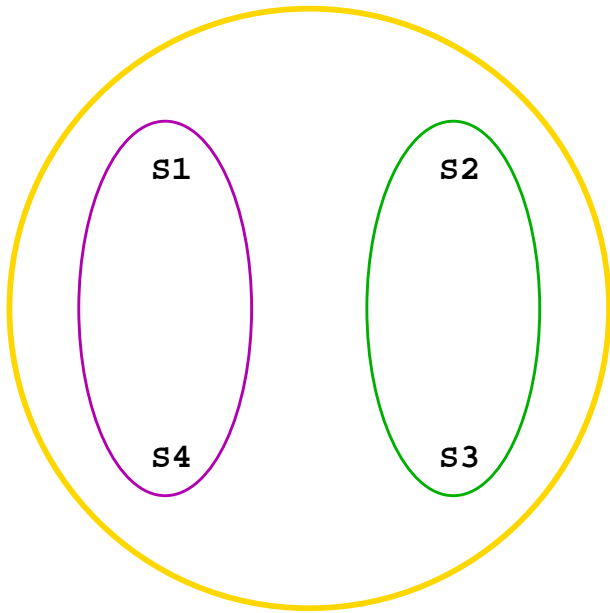
Tableau des scores d'alignement:

	s_1	s_2	s_3	s_4
s_1		2	0	17
s_2	2		14	0
s_3	0	14		-1
s_4	17	0	-1	

n séquences
 ↓
 $n(n - 1)/2$ calculs

Étape 2 : construction de l'arbre guide

Arbre obtenu avec l'algorithme de Neighbor-Joining



Les séquences sont regroupées suivant leur similarité à partir de la matrice des scores 2 à 2.

dendogramme

Étape 3 : *construction de l'alignement multiple final*

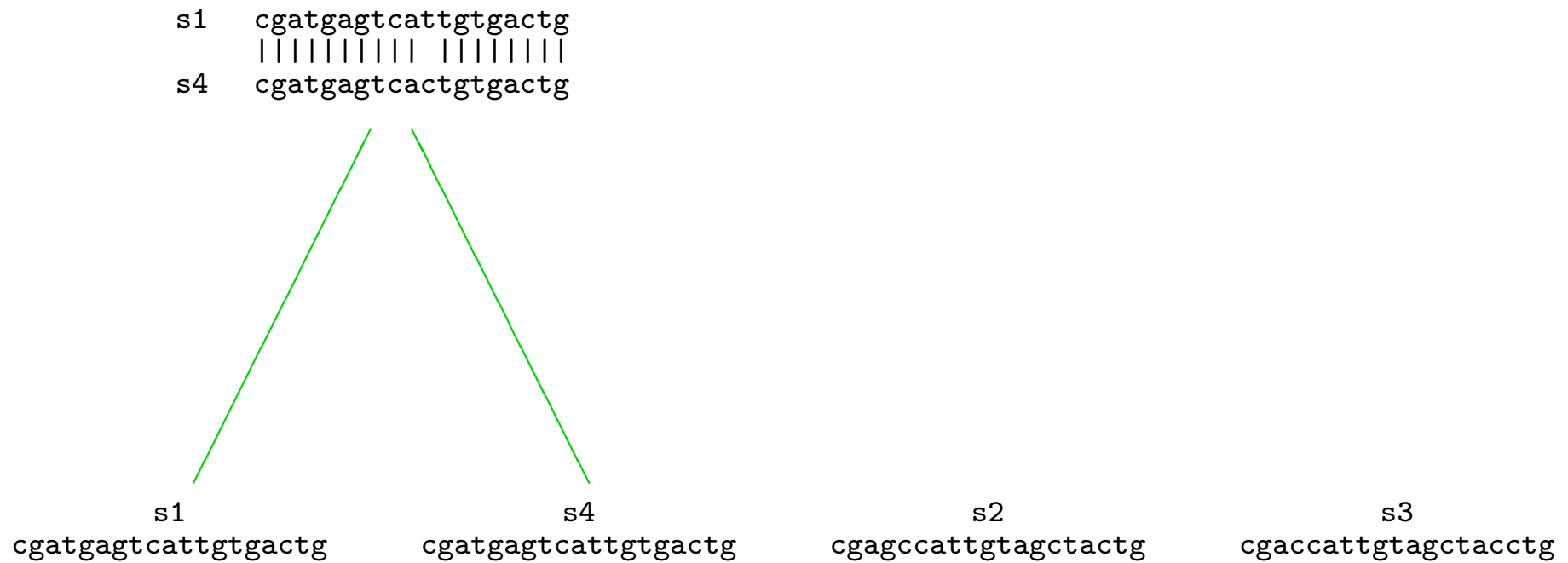
s1
cgatgagtcattgtgactg

s4
cgatgagtcattgtgactg

s2
cgagccattgtagctactg

s3
cgaccattgtagctacctg

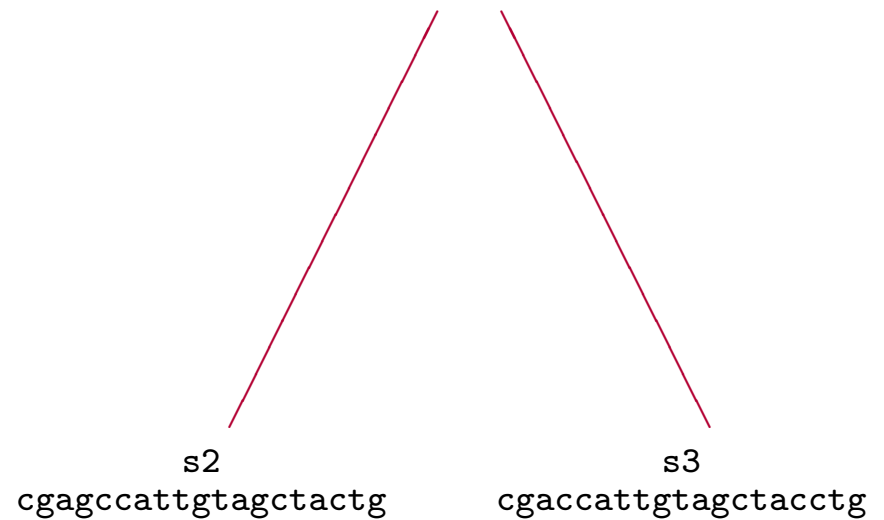
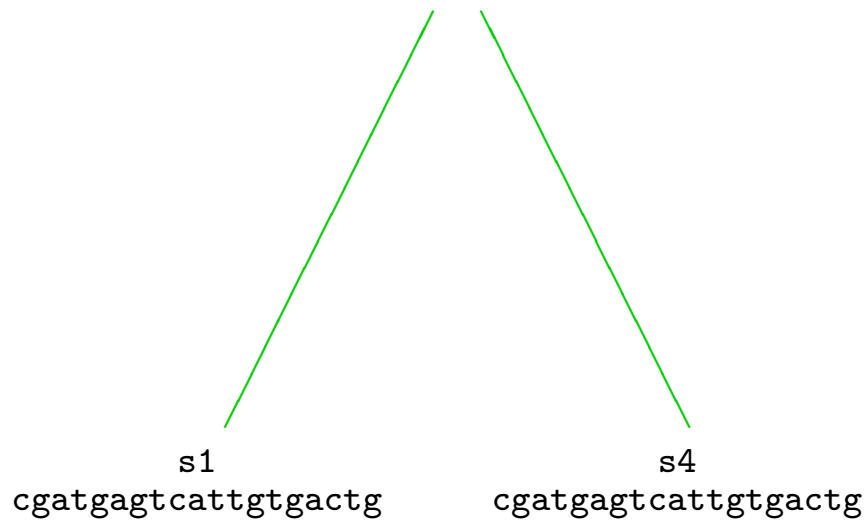
Étape 3 : construction de l'alignement multiple final



Étape 3 : construction de l'alignement multiple final

s1 cgatgagtcattgtgactg
|||||
s4 cgatgagtcactgtgactg

s2 cgagccattgtagcta-ctg
|||
s3 cga-ccattgtagctacctg



Étape 3 : construction de l'alignement multiple final

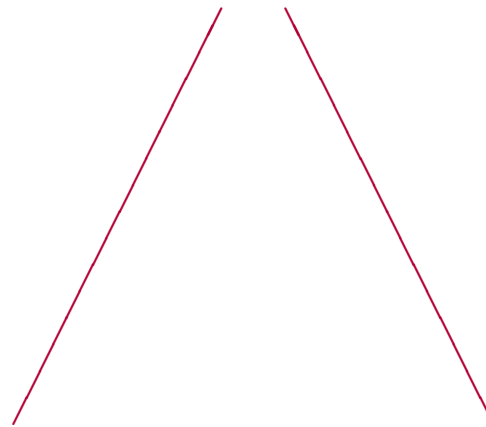
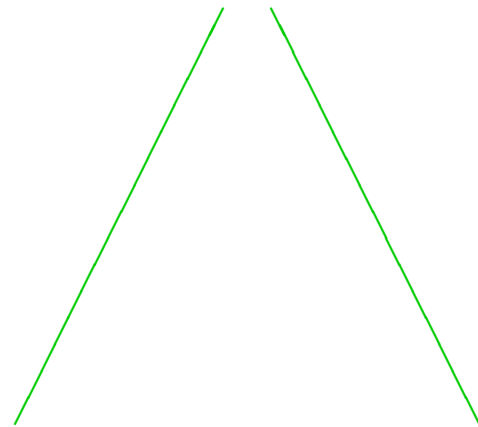
"Once a gap, always a gap."

```
s2  cga---gccattgtagctac-tg
s3  cga----ccattgtagctacctg
s1  cgatgagtcattgt-g--ac-tg
s4  cgatgagtcactgt-g--ac-tg
```



```
s1  cgatgagtcattgtgactg
    |||||
s4  cgatgagtcactgtgactg
```

```
s2  cgagccattgtagcta-ctg
    |||
s3  cga-ccattgtagctacctg
```



```
s1  cgatgagtcattgtgactg
s4  cgatgagtcattgtgactg
```

```
s2  cgagccattgtagctactg
s3  cgaccattgtagctacctg
```

Clustalw est optimisé pour les protéines

- ▷ Pondération des séquences en fonction de leur sur- ou sous- représentation
- ▷ Adaptation des matrices de similarité au fil de l'algorithme en fonction de la divergence des séquences à aligner

BLOSUM 80 pour aligner les séquences proches,

BLOSUM 50 pour aligner des séquences distantes, par exemple.

- ▷ Pénalités de gaps spécifiques à chaque résidu.

Par exemple, les glycines sont davantage susceptibles 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

Paramètres de Clustal

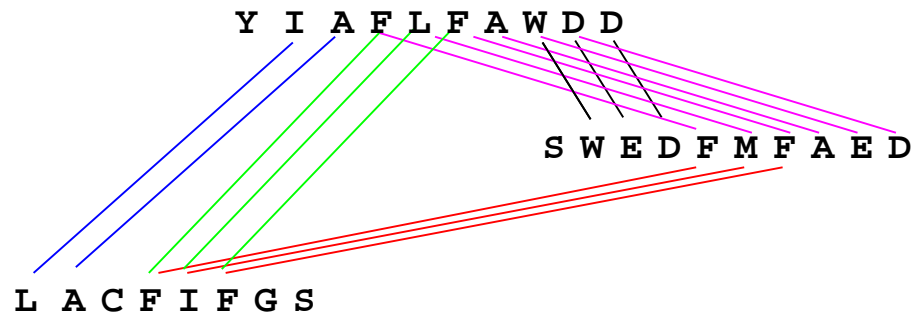
- ▷ *Slow/Fast* : qualité des alignements 2 à 2
- ▷ Matrice de similarité (PAM, BLOSUM, Gonnet)
- ▷ Pénalités de gaps :
 - Ouverture et d'extension de gaps
 - Distance de voisinage entre deux gaps
 - Gaps hydrophiles
 - Ouvertures de gaps spécifiques

Dialign2

Morgenstern *et al.* - 1999

Construire un alignement à partir des diagonales du dot plot

Étape 1: Détection des diagonales dans les paires de séquences



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

- Pas de croisements
- Pas de chevauchements
- Score maximal

```
y I A - F L F A W D d
- L A c F I F g s - -
s w e d F M F A E D -
```

Exemple (C. Notre-Dame)

GARFIELD THE LAST FAT CAT
GARFIELD THE FAT CAT
GARFIELD THE VERY FAST CAT
THE FAT CAT

Alignement fourni par Clustal

```
seq1      GARFIELDTHELASTFA-TCAT
seq2      ----GARFIELDTHEFA-TCAT
seq3      GARFIELDTHEVERYFASTCAT
seq4      -----THEFA-TCAT
```

Alignement fourni par Dialign2

```
seq1      GARFIELD THE LAST FA-T CAT
seq2      GARFIELD THE ---- FA-T CAT
seq3      GARFIELD THE VERY FAST CAT
seq4      ----- THE ---- FA-T CAT
```



Quelle méthode utiliser ?

Cela dépend du type de séquences à aligner . . .

BaliBASE : base de données d'alignements multiples et de benchmarks

- ▷ plus de 150 familles de protéines
- ▷ alignements basés sur la structure secondaire

Référence 1 séquences équidistantes avec différents niveaux de conservation

Référence 2 protéines homologues + 1 séquence orpheline

Référence 3 sous-groupes avec moins de 25% d'identité entre les groupes

Référence 4 extensions N/C-terminales

Référence 5 insertions internes

Référence 6 répétitions internes

Référence 7 protéines transmembranaires

Référence 8 permutations de domaines

Réf. 1, 2 et 3: préférer **Clustal** à **Dialign2**

Réf. 4 et 5: préférer **Dialign2** à **Clustal**

- ▷ Plus les séquences sont divergentes, moins le résultat est fiable
 - Quand le taux d'identité est supérieur à 35%, toutes les méthodes sont satisfaisantes
 - Alignements corrects à plus de 90%*
 - twilight zone : 10-20 % identité
 - Aucune méthode n'assure un alignement avec plus de 50% de correction*

- ▷ **Clustal** a tendance à autoriser moins de gaps que **Dialign2**

- ▷ Similarité locale : **Dialign2**

- ▷ Similarité globale : **Clustal**

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

Exemple : domaine SH3

SH3 (Src homology 3) domains are often indicative of a protein involved in signal transduction related to cytoskeletal organization. The SH3 domain has a characteristic 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.

Prosite PS50002

Séquences à aligner	longueur
=====	=====
1aboA P00520	57
1ycaB P04637	60
1pht P27986	80
1ihvA P00383	49
1vie P12497	51

- ▷ Séquences courtes
- ▷ Similarité faible (< 25%) et diffuse

SH3 - Véritable Alignement

Basé sur l'alignement des éléments de structure secondaire

```
1aboA  -NLFVALYDfvasgdntlsitkGEKLRVLgynhn-----
1ycsB  kGVIYALWDyepqnddelpmkeGDCMTIIhrede-----
1pht   gYQYRALYDykkereedidlhlGDILTVNkgs lvalgfsd
1ihvA  -NFRVYYRDSrd-----pvwkGPAKLLWkg-----
1vie   -drvrkksga-----awqGQIVGWYctnlt-----

1aboA  -----gEWCEAQt--kngqGWVPSNYITPVN-----
1ycsB  -----deiEWWWARl--ndkeGYVPRNLLGLYP-----
1pht   gqearpeeiGWLNGYnettgerGDFPGTYVEYIGrkkisp
1ihvA  -----eGAVVIQd--nsdiKVVPRRKAKIIRd-----
1vie   -----peGYAVESeahpgsvQIYPVAALERIN-----
```

SH3 - Alignement fourni par Clustal

```
1aboA  -NLFV-ALYDFVASGDNTLSITKGEKLRV-----LGYNHNG
1ycsB  KGVIIY-ALWDYEPQNDDELPMKEGDCMTI-----IHREDED
1pht   -GYQYRALYDYKKEREEDIDLHLGDILTVNKGSLVALGFSDGQ
1ihvA  -----NFRVYYRDSRD--PVWKGPAKLL-----WKGEG
1vie   -----DRVRKKSG--AAWQGQIVGW-----YCTNL
```

```
1aboA  -----EWCEA--QTKNGQGWVPSNYITPVN-----
1ycsB  EI-----EWWWA--RLNDKEGYVPRNLLGLYP-----
1pht   EARPEEIGWLNGYNETTGERGDFPGTYVEYIGRKKISP
1ihvA  -----AVVIQ---DNSDIKVVPRRKAKIIRD-----
1vie   TP----EGYAVESEAHPGSVQIYPVAALERIN-----
```

SH3 - Aligment fourni par Dialign2

```
1aboA  n-LFVALYDFVASGDNTLSITKGEKLRVL-----  
1ycsB  kgVIYALWDYEPQNDDELPMKEGDCMTIIhr---EDEDEI-----  
1pht   gyQYRALYDYKKEREEDIDLHLGDILTVNKGSLVALGFSDgqearpeei  
1ihvA  --NFRV---YYRDSRDPVWKGPAKLLWKGEGAVVIQDNSDI-----  
1vie   -----DRVRKKSGaa-W-----QGQI-----
```

```
1aboA  ----GYNhngEWCEAQTKNGQGwV-----PSNYItp-----VN  
1ycsB  -----EWWARLNDKEGYV-----PRNLLgLYP-----  
1pht   gwlnGYN-----ETTGERGDF-----PGTYV-EYigRKKIsp--  
1ihvA  -----Kv-----V-----PRr-----KAKIIRd-  
1vie   -----VGWYCTNLTPEGYAveseahPGSVQ-IYPv-AALERIN
```

Exemple : cinq protéines avec un domaine *helix - loop - helix*

Séquences à aligner:

=====

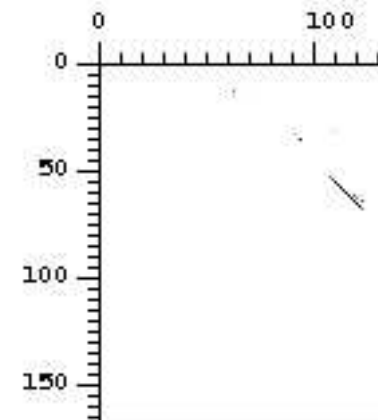
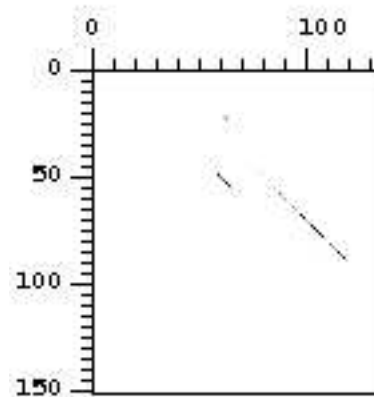
longueur:

=====

- | | | |
|----|------------|-----|
| 1) | HEN1-Human | 133 |
| 2) | CBF1-Yeast | 351 |
| 3) | HES5-Mouse | 167 |
| 4) | INO4-Yeast | 151 |
| 5) | ESC1-Yeast | 413 |

▷ longueurs dissemblables

▷ similarité locale



helix-loop-helix, alignment Clustal

-----MMLNSDTMELD-----LPPTHSETESG-----FSDCGGG
--MNSLANNNKLSTEDDEIHSARKRGYNEEQNYSEARKKQRDQGLLSQESNDGNIDSALLSEGATLKGTQSQYESG-----LTSNKDE
MSSYALPSMQPTPTSSIPLRQMSQPPTSAPSNSASSTPYSPQQVPLTHNSYPLSTPSSFQHGQTRLPPINCLAEPFNRQPWHSNSAAP
-----MAPSTVAVEMLSPKEKNRLRKPVVEKMRRDR-----INSSIEQ
-----MTNDIKEIQTIQPGLSEIKEIKGELANVKKR-----

AGPD-----GAGPGG-----
KGSDDDEDASVAEAAVAATVNYTDLIQQQE-----DSSDAHTSNQTNANGEHKDSLNGERAITPSNEGVPNTSLEGMTSSPMEST
ASSSPTSATLSTA AHPVHTNAAQVAGSSSSSYVYSVPPTNSTTSQASAKHSAPHRSSQFQSTTLTPSTTDSSTDVSSSDSVSTSASSS
LKLL-----

-----PGGGQARGPEPEPGRKD-----LQHLSREERRRRRAT-----AKYRTA-----
QQSKNDMLIPLAEHDRGPEHQDDEDNDDADID-----LKKDISMQPGRGRKPTTLATTDEWKKQRKDS-----
NASNTVSVTSPASSATPLPNQPSQQQFLVSKNDAFTTFVHSVHNTPMQQSMYVPQQQTSHSSGASYQNESANPPVQSPMQYSYSQGGP
-----LEQEFARHQPNKLEKAD-----ILEMAVSYLKHSKAFAAAAGPKSLHQDYSEG-----
-----KRRSKKINKLTDGQIR-----INHVSSEKKRRELERAI FDELVAVVPDLQPQ-----

-----HATRERIRVEAFNLAFA--ELRKLPTLPP-----DKKLSKIEILR
-----HKEVERRRRRENINTAIN--VLSDLLPVRESSKAAILARAAEYIQLKETDEANIEKWTLQKLLSEQNASQ
FSYPQHKNQSFSA SPIDPSMSYVYRAPESFSSINANVPYGRNEYLRVTS LVPNQPEYTG PYTRNPELRTSHKLAERKRRKEIKELFDDLKDA
-----YSWCLQEAVQFLTLHAASDTQMKLLYHFQRP-----APAAPAKEPPA
-----ESRSELI IYLSLSYLSWLYERNEKLR-----KQIIAKHEAKT

helix-loop-helix, alignment Dialign2

```
mml-----  
m-----NSLANNNKLS  
MA-----  
MT-----  
mssyalpsmqptptssiplrqmsqpttsapsnsasstpypqqvplthnsyplstpssfqhggqtrlppinclaePFNRPQPWHSNSAAPASSSPTSATLS  
  
-----NSDTMELD-----LPPTHSETESGFSDCGGGAGPDgagpgggpgggqarg-----  
TEDEEIHSAKRGYNEEQNYsearkkqrdqgllsqesndgnidsallsegatLKGTQSQYESGLTSNKDEKGSDDedasvaeaavaatvnytdliqgQED  
-----  
TAAHPVHTNAAQVAGSSSSYVYS-----VPPTNSTTSQAsakhsavphrssfqqsttltptt-----DSS  
  
-----PEPGEPRK-----  
SSDAHTSNQTNANGEHKDSLNGERAITPSNEGPK-----NTSLEGMTSSPMESTQQSKNdmliplaehdrg-----  
-----  
STDVSSSDSVSTASSSNASNTVSVTSPASSSATPLPNQPSQQqflvskndafttfvhsvhNTPMQQSMYVPQQQTSHSSGasyqnesanppvqspmqys
```

-----DLQHL---SREERRRRRATA-----K
-----PEHQqddednddadidlkkdismqpgrrgrkPTTLAtt dew-KKQR-----
-----PSTVAVEMLSPEKN-----
-----NDIKEIQTIQPGLSEIKEIKGELANVKKR---KRRSKKINKLTDG-----Q
ysqgqpf syPQHK-----NQSFSASPIDPSMSYVYRAPESFSSINANvpyGRNEYLRrvtslvpnqpeytgpytrnpE

YRTAHATRERIRVEAFNLAFaelRkLLPTL----PPDKKLSKIEILRLAICYISYLNHVldv-----
-KDSHKEVERRRRENINTAINVLSDLLP-V----RESSKAA---ILARAAEYIQLKkETDEanieKWTLQKLLSEQNASQLASANEKLQEELGNaykeie
-RLRKPVVEKMRRDRINSSIEQLKLLLeqefarhQPNSKLEKADILEMAVSYLKHSKAFAA----Aag-----P
IRINHVSSSEKKRRELERAI FDELVAVVPDL----QPQESRSELI IYLKSLSYLSWLYERNE---KLRKQIIAKHEAKTGSSSSSDPVQE QNGNirdlvP
LRTSHKLAERKRRKEIKELFDDLKDALP-L----DKSTKSSKWGLLTRAIQYIEQLKSEQV----ALEAYVKSLEEnmqsnkevtkgt-----

ymkrv1r-----KEGIEYEDMHTHkkqenerkstrsdnphea-----
KSLHQDYSEGYSwclQEAVQFLTLHAasdtqmkllyhfqrppapaapakeppapgaapqparssakaaaaavstsrqpacglwrpw
KELIWELGDGQSgq-----
