

EACS: Architecture, Résultats et discussions

Présenté par
Khalid BENABDESLEM
kbenabde@ibcp.fr

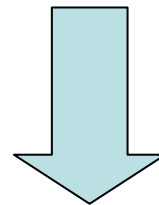
En collaboration avec
Christophe Ceourjon
Emmanuel Bettler

Projet
GENOTO3D - ACI: Masse de données

Extraction automatique des cœurs structuraux

Outils:

- Une méthode de classification de structures (CATH)
- Une méthode d'alignement structural (CE)
- Une méthode de classification automatique (CAH)

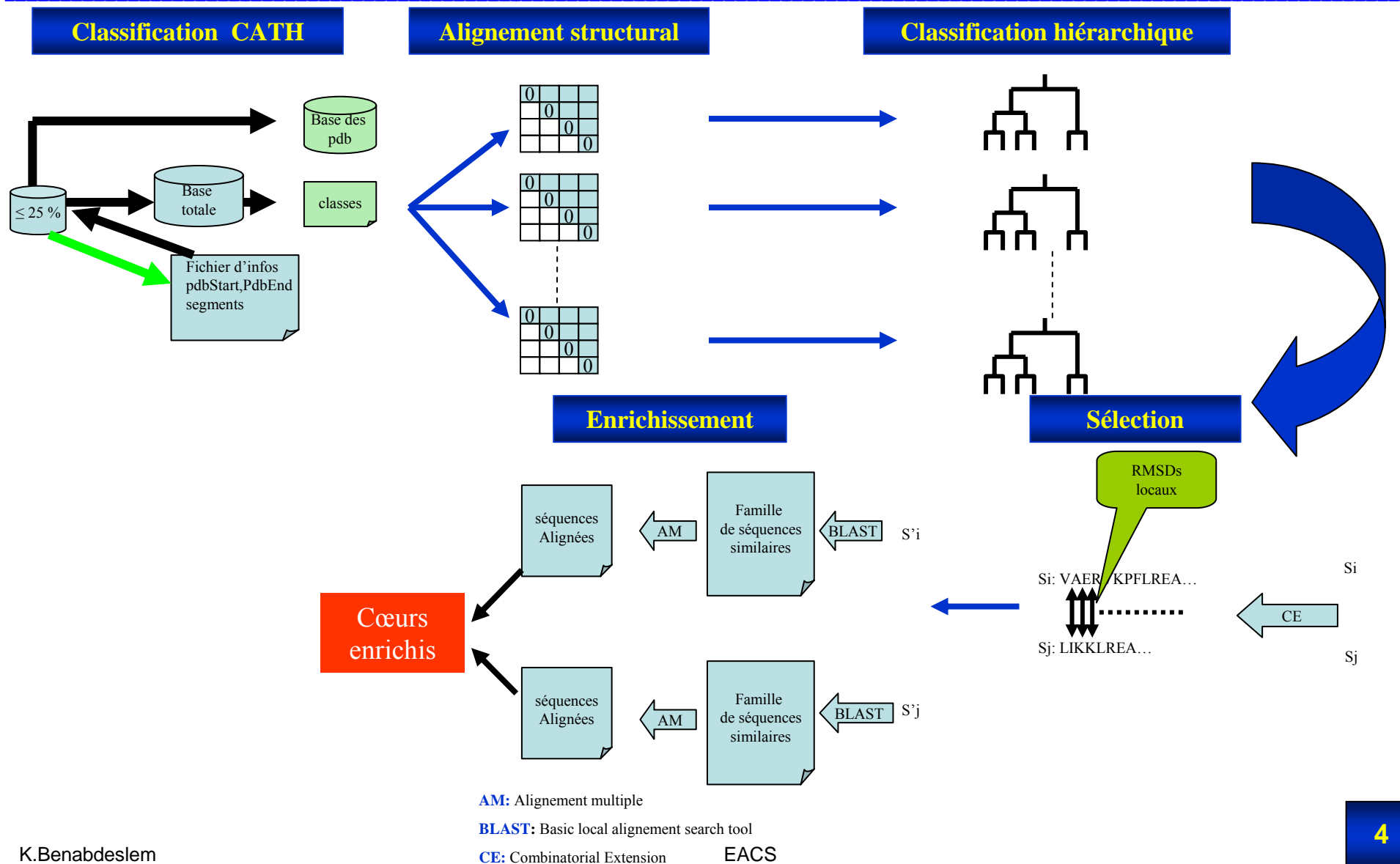


EACS

EACS: Fonctionnalités

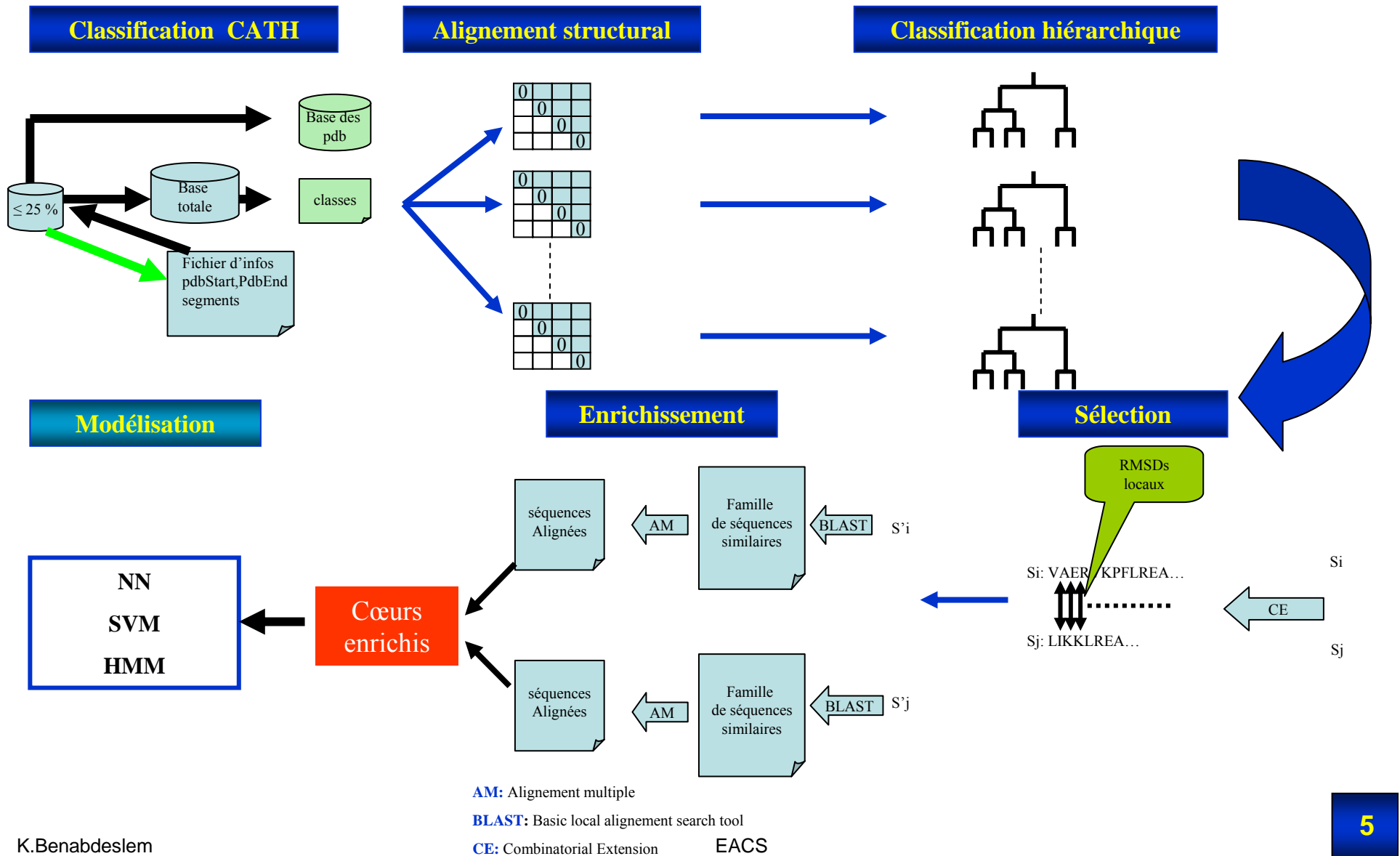
- Des cœurs structuraux à différents niveaux
- Intersection significative des structures d'une même famille
- Prototypage des classes de protéines
- Espace de données «significatif »
- Prétraitement pour la reconnaissance de repliements ou modélisation 3D

EACS: Architecture générale



AM: Alignement multiple
BLAST: Basic local alignment search tool
CE: Combinatorial Extension

EACS: Architecture générale



EACS: Politique logicielle

- Version A

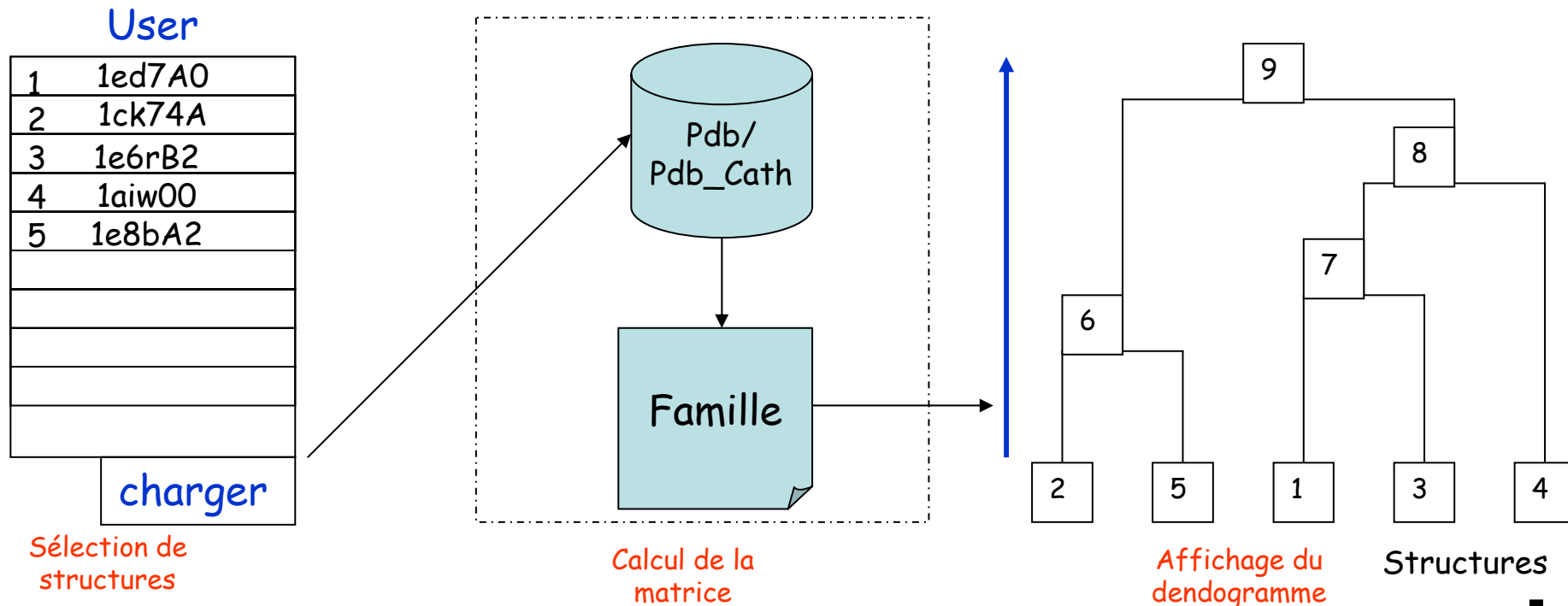
- **A.1:** des structures à un seul segment
- **A.2:** des structures à plusieurs segments + enrichissement
- **A.3:** des structures à plusieurs modèles

- Version B

- **B.1:** Mise à jour de structures « utilisateur » + recalcul
- **B.2:** Adaptation du modèle aux structures « utilisateur »

EACS:

Exemple d'application (A.1)



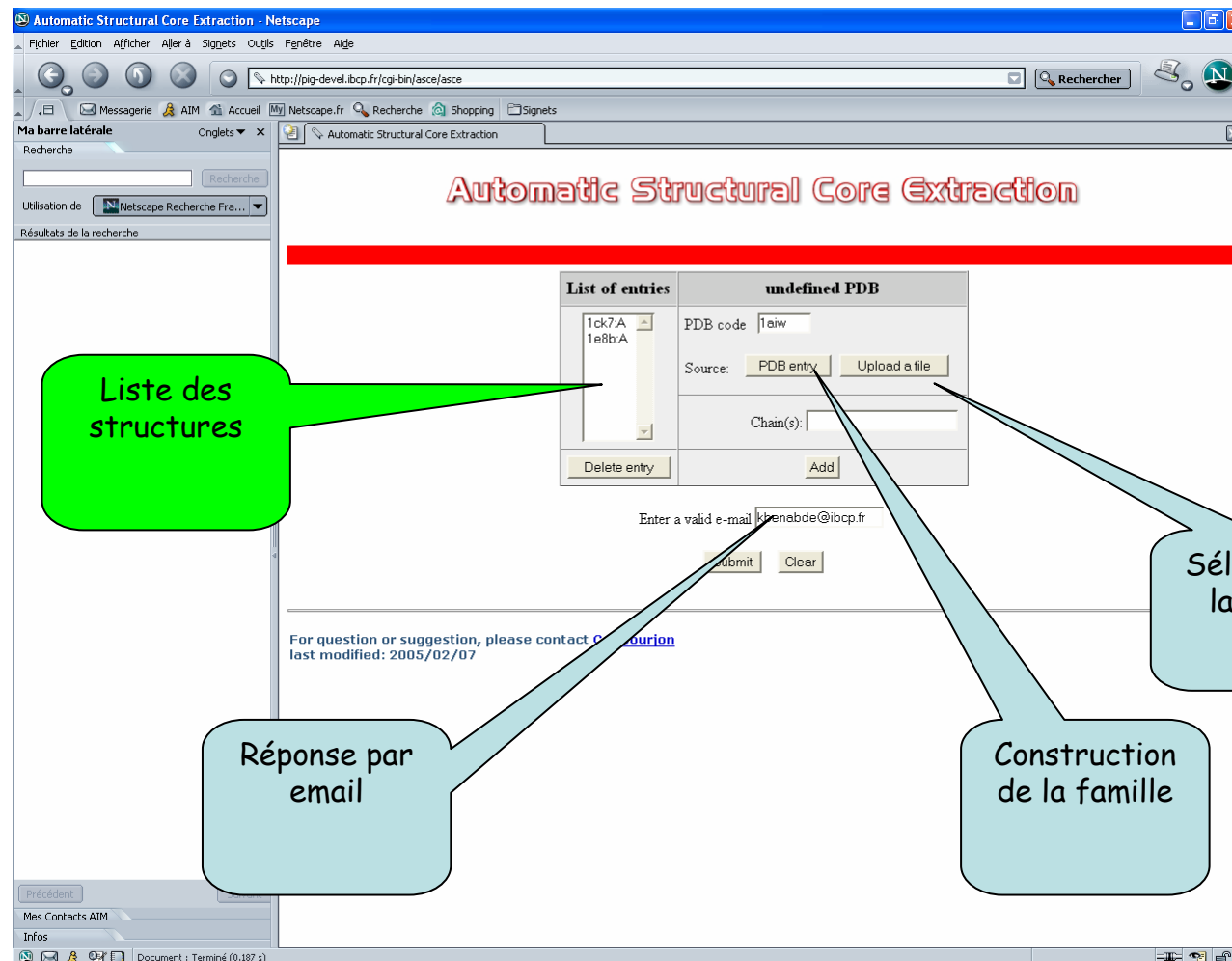
- 1^{ère} composante du cœur
- 2^{ème} composante du cœur
- Pourcentage de préservation
- Pourcentage de population
- Structures secondaires des cœurs
- Informations sur les cœurs

User

Nombre de coeurs

EACS: V A.1

Vue Web



The screenshot shows a Netscape browser window titled "Automatic Structural Core Extraction - Netscape". The address bar contains the URL `http://pbig-devel.ibcp.fr/cgi-bin/asce/asce`. The main content area features the title "Automatic Structural Core Extraction" in red text. Below the title is a form with two main sections: "List of entries" and "undefined PDB".

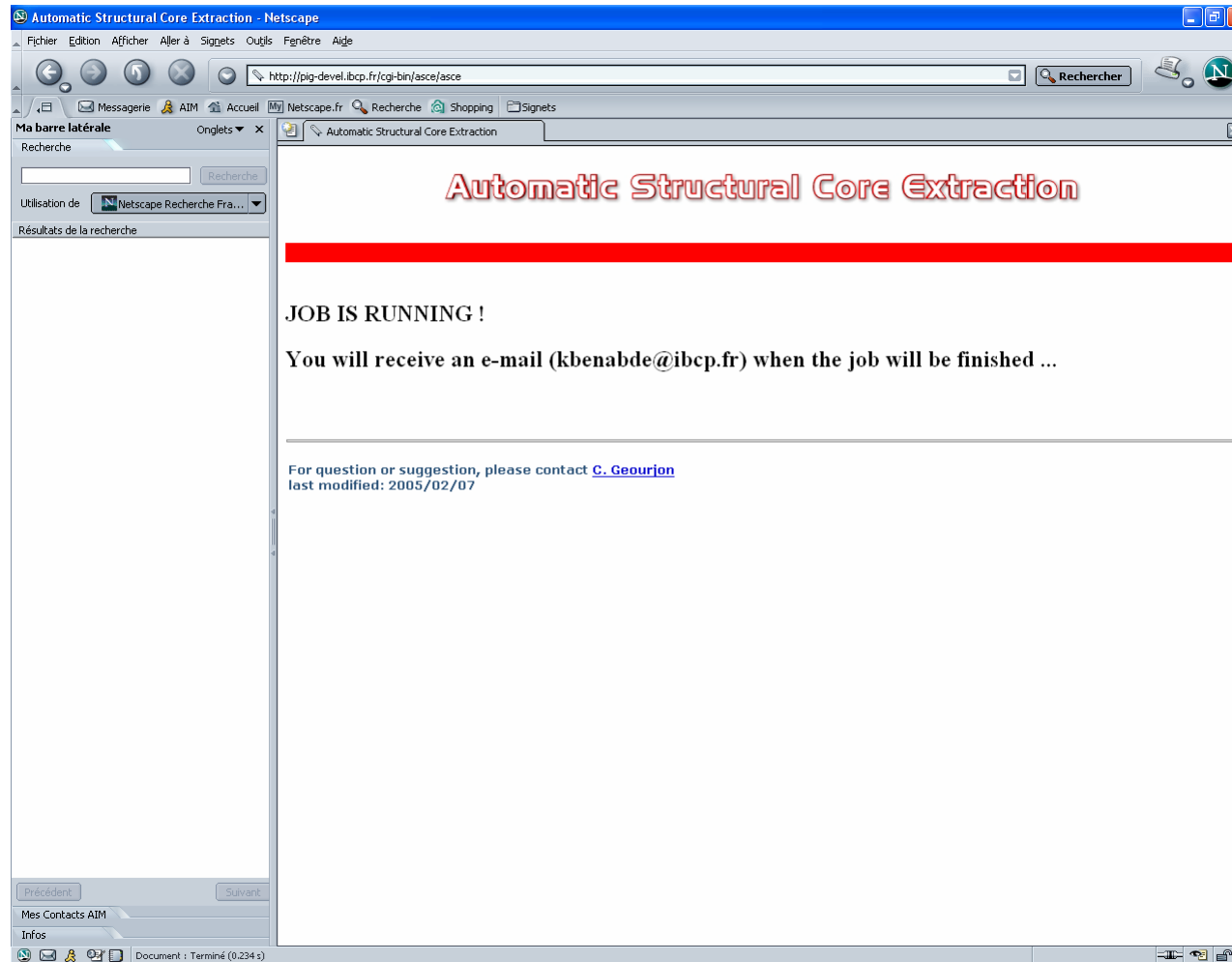
The "List of entries" section contains a list box with the entries "1ck7:A" and "1e8b:A", and a "Delete entry" button. The "undefined PDB" section includes a "PDB code" field with the value "1eiw", a "Source" field with a dropdown menu set to "PDB entry" and an "Upload a file" button, and a "Chain(s)" field. An "Add" button is located at the bottom of this section.

Below the form is an email input field with the text "Enter a valid e-mail" and the address "kbenabde@ibcp.fr", along with "Submit" and "Clear" buttons. At the bottom of the page, there is a footer that reads: "For question or suggestion, please contact [C. Jourjon](#) last modified: 2005/02/07".

Four callout boxes are overlaid on the screenshot:

- A green callout box labeled "Liste des structures" points to the list of entries.
- A light blue callout box labeled "Sélection de la famille" points to the "PDB entry" dropdown menu.
- A light blue callout box labeled "Construction de la famille" points to the "Add" button.
- A light blue callout box labeled "Réponse par email" points to the email input field.

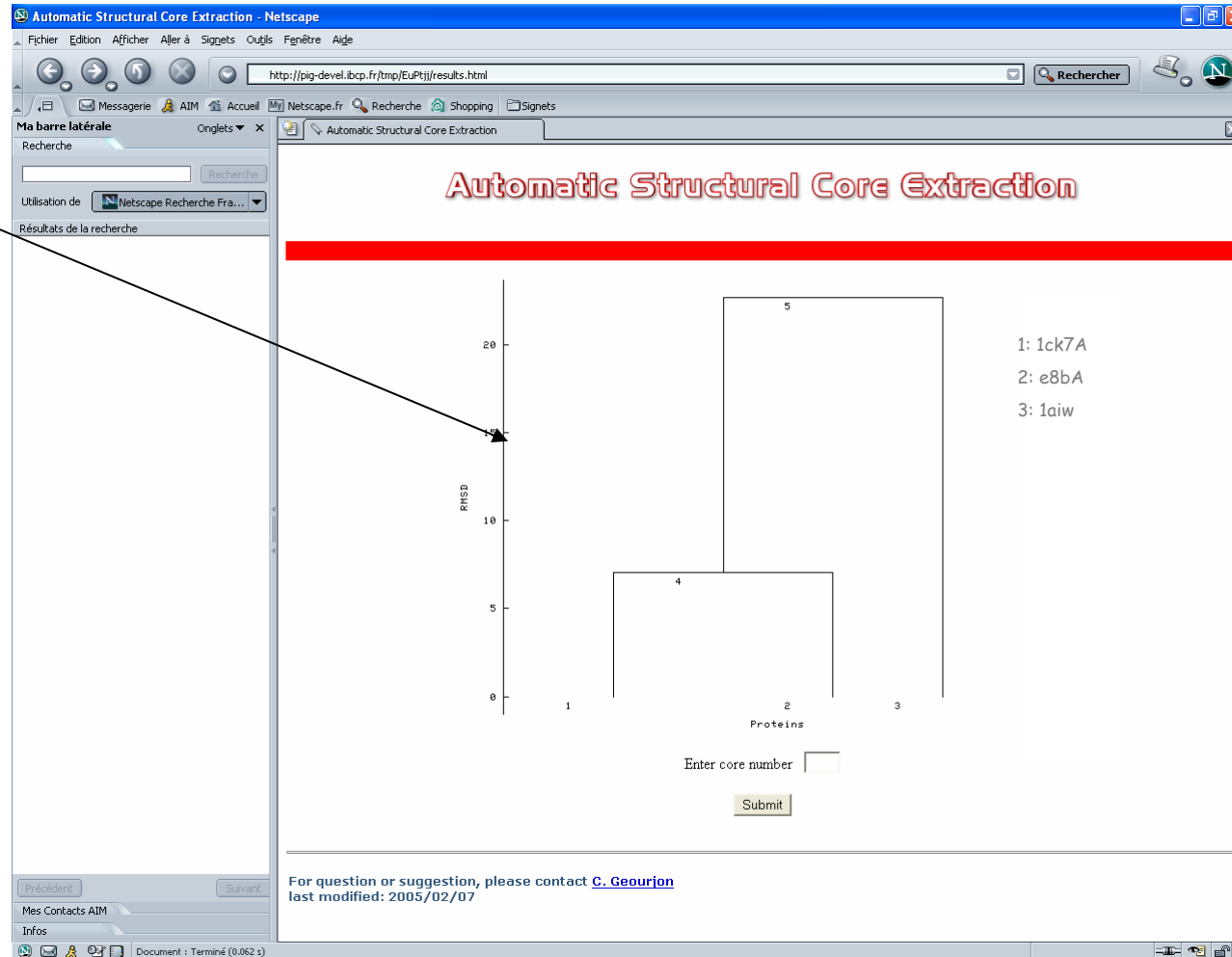
EACS: V A.1 Vue Web



EACS: V A.1 Vue Web

User

URL



Une page contenant



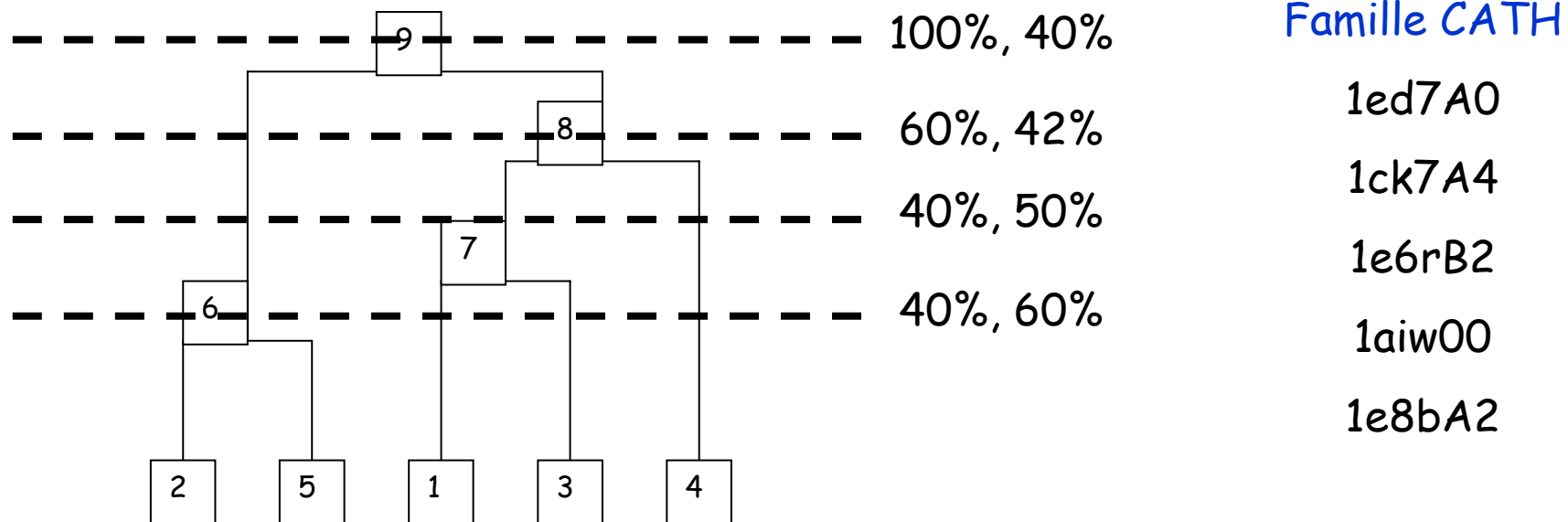
Les composantes des cœurs

Les séquences des cœurs avec coloration des structures secondaires

La représentation 3D des cœurs

Le pourcentage de préservation moyen des résidus dans les cœurs

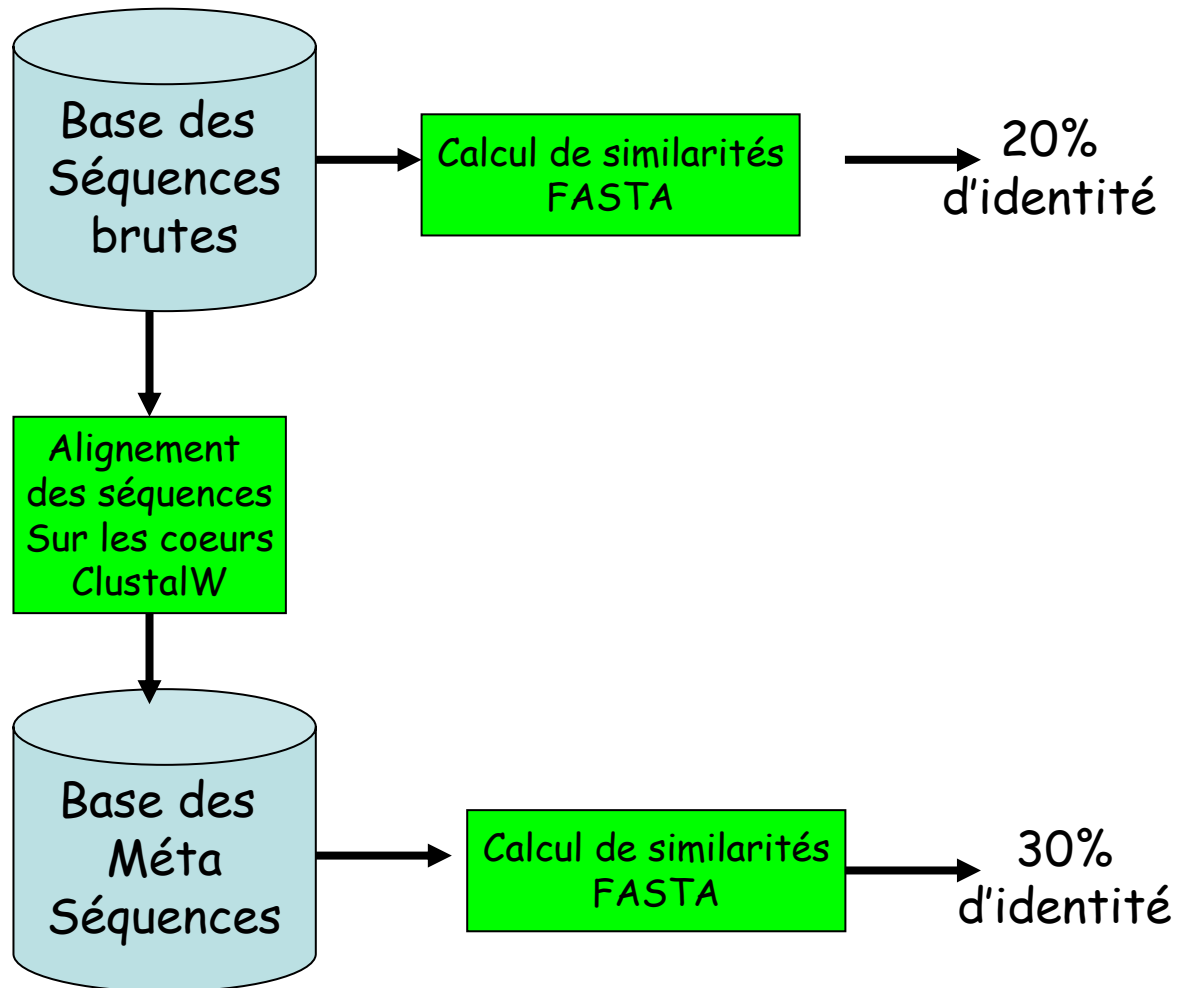
Le pourcentage de population des cœurs



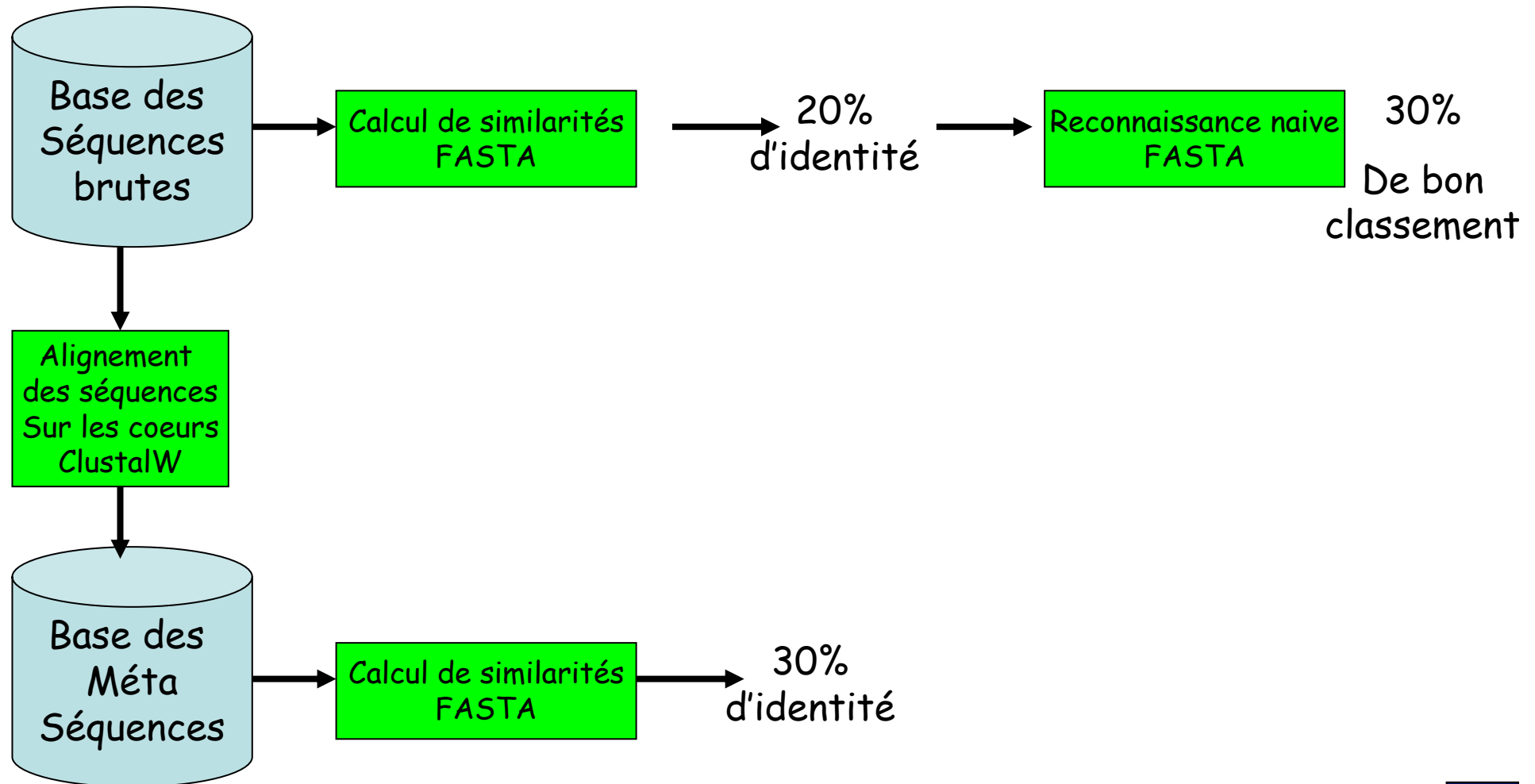
190 structures par 85 familles CATH (60% à 2 structures)

Pourcentage moyen de couverture de résidus : 60% sur un
 pourcentage moyen de population de 64 %

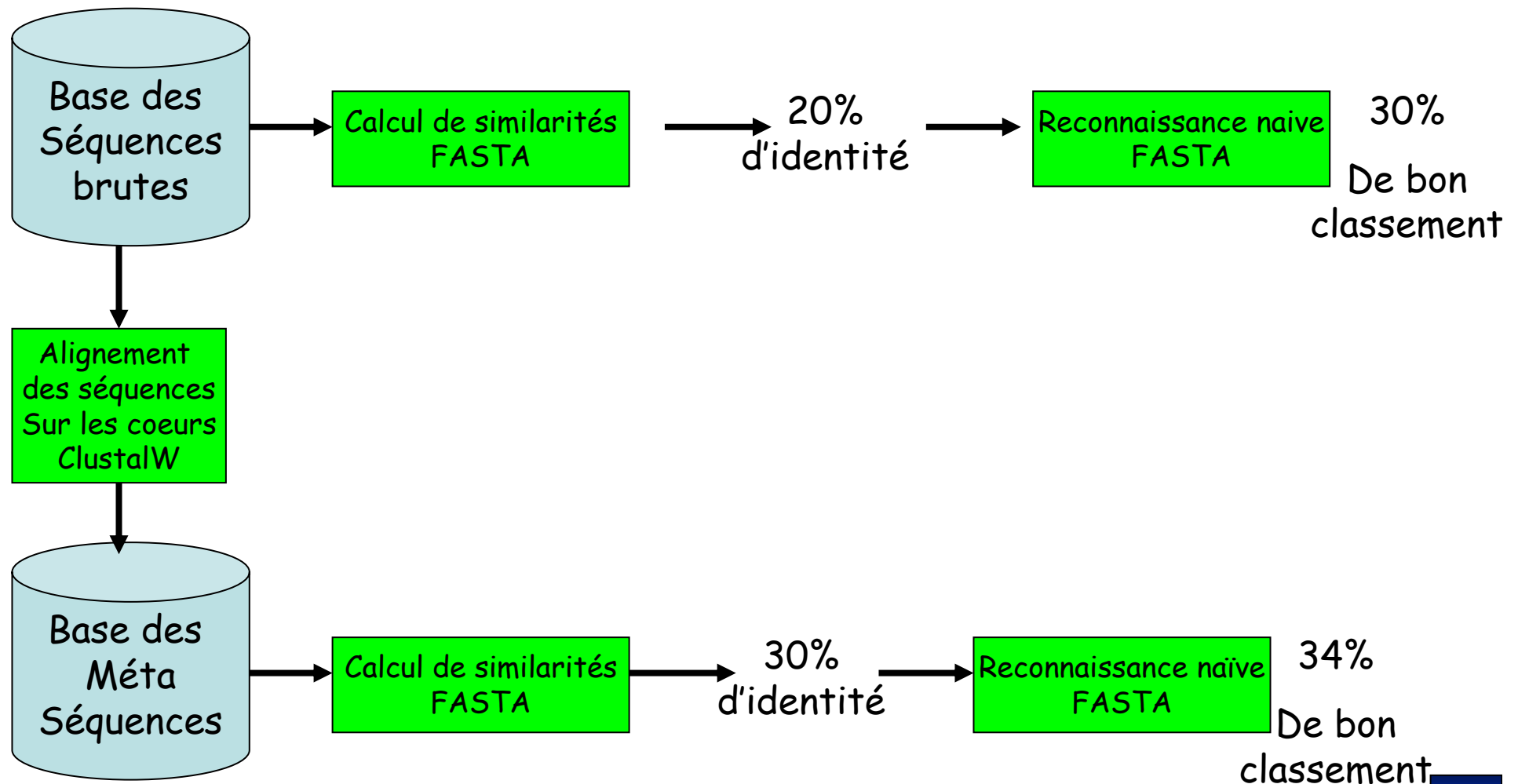




Base alignée sur les cœurs



Base alignée sur les cœurs



Reconnaissance des cœurs par NN

Seq

AIVVDDSVFSPSYVPKRYTKDQIFDILLDRAKAGLAEG.....

Alignement sur les cœurs

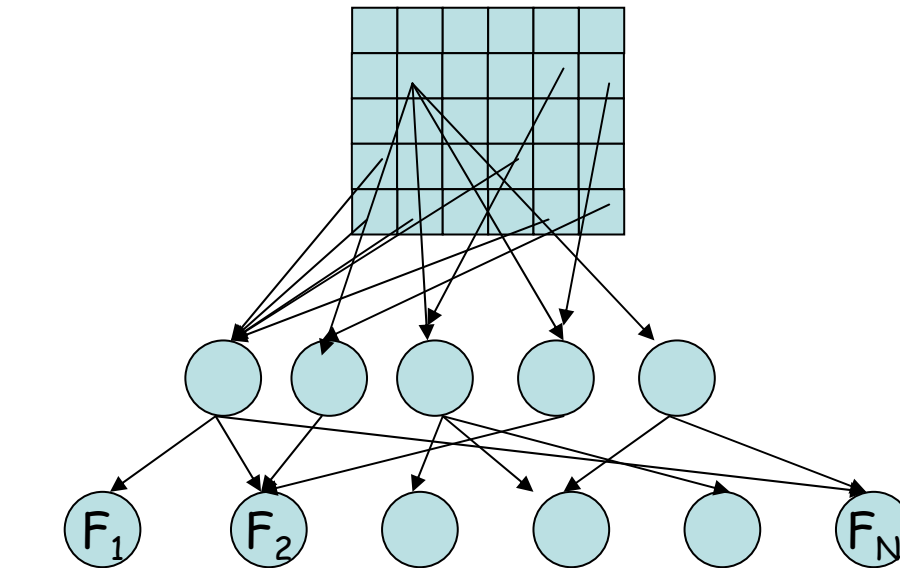
M-Seq

AIVVDDXXXSVFSPSYVPXXKRYTKDQIFDILLDRAKAGLAEG.....

Codage



Matrice COVseq modélisant la séquence



Fonction de coût: Distance matricielle:
(Frobenius)

$$fd(COVx, W) = \sum_i \sqrt{\sum_j (COV_{ij} - w_{ij})^2}$$

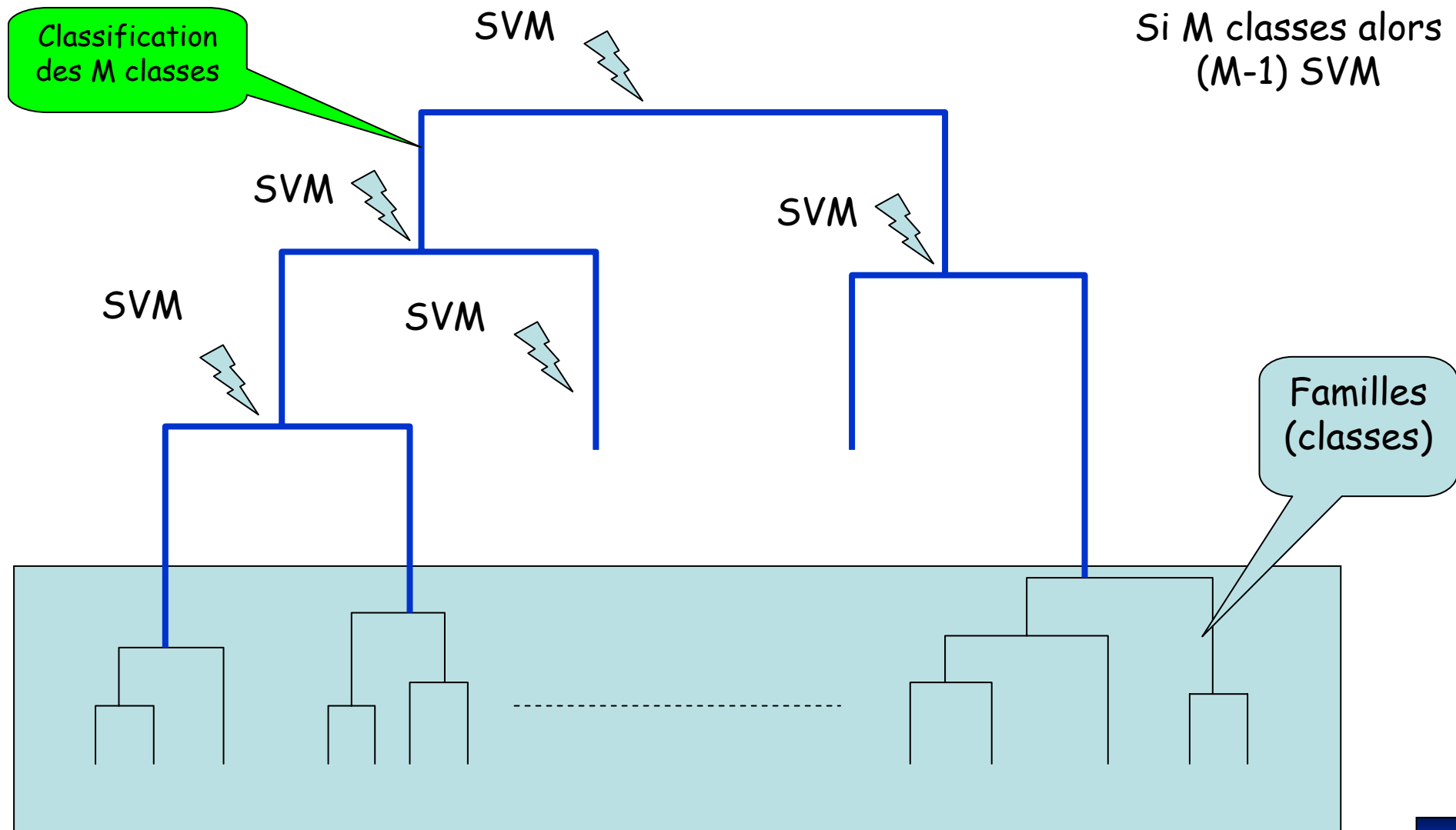
Reconnaissance des cœurs par NN: Résultats

Une base de 80 exemples sur 21 familles (à p-structures)

75% d'apprentissage et 25 % de test

~70 % de bonne reconnaissance séquence - famille

Ambitions: CH-SVM



Bilan et Perspectives

- Outil EACS (Application Web + papier CAP'05)
- MLP basé sur distance matricielle sur des méta - séquences modélisées (Modélisation + Résultats)
- Une idée préliminaire sur les CH-SVM = CAH + SVM = M-SVM

- Améliorer EACS (Enrichissement + Évolutivité + d'autres outils d'alignement)
- Améliorer les résultats du MLP après enrichissement de la base
- Une métrique de regroupement des classes (structuralement) pour le CH-SVM