



HAL
open science

Site de calculs du paramètre de solubilité moléculaire

Olivier Garrigues

► **To cite this version:**

Olivier Garrigues. Site de calculs du paramètre de solubilité moléculaire. Génie logiciel [cs.SE]. 2010. dumas-00524180

HAL Id: dumas-00524180

<https://dumas.ccsd.cnrs.fr/dumas-00524180v1>

Submitted on 7 Oct 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CONSERVATOIRE NATIONAL DES ARTS ET METIERS
CENTRE REGIONAL ASSOCIE DE LYON

MEMOIRE

présenté en vue d'obtenir
le **DIPLOME d'INGENIEUR CNAM**

SPECIALITE : INFORMATIQUE

option : Système d'information

par

GARRIGUES Olivier

Site de calculs du paramètre de solubilité moléculaire

(Projet C.P.S.M)

Soutenu le : 25 juin 2010

JURY

PRESIDENT : CHRISTOPHE PICOULEAU

MEMBRES: BERTRAND DAVID

CLAUDE GENIER

ENCADRANTS : HAMID ELAISSARI

CLAUDIA ELIZABETH MORA-HUERTAS

Projet C.P.S.M.

Table des matières

<u>I Introduction</u>	9
<u>II Contexte</u>	11
<u>II.1 Mon travail</u>	11
<u>II.2 Rôle dans l'université</u>	13
<u>III État des lieux</u>	15
<u>III.1 Personnes et besoins</u>	15
<u>III.2 Système informatique</u>	16
<u>IV Analyse et solutions envisageables</u>	23
<u>IV.1 Expression des besoins</u>	23
<u>IV.2 Analyse des besoins</u>	24
<u>IV.3 Étude de l'existant</u>	29
<u>IV.4 Étude des besoins</u>	30
<u>IV.5 Modèle conceptuel de données</u>	32
<u>IV.6 Paramètre de solubilité</u>	39
<u>IV.7 Différentes méthodes de calculs</u>	42
<u>IV.8 Respect des contraintes du cahier des charges</u>	47
<u>IV.9 Choix des plateformes</u>	51
<u>V Organisation du développement</u>	57
<u>V.1 Organisation administrative</u>	57
<u>V.2 Organisation humaine</u>	57
<u>V.3 Organisation de la réalisation</u>	58
<u>V.4 La gestion du projet</u>	60
<u>VI Réalisation</u>	67
<u>VI.1 Solutions de programmation</u>	67

Projet C.P.S.M.

<u>VI.2 base de données</u>	70
<u>VI.3 Les traitements</u>	77
<u>VI.4 Plan du site</u>	88
<u>VI.5 Tests unitaires</u>	98
<u>VII Contraintes rencontrées</u>	103
<u>VII.1 Contraintes internes</u>	103
<u>VII.2 Contraintes externes</u>	104
<u>VIII Bilan et perspectives</u>	109
<u>IX Annexe A : Bibliographie</u>	111
<u>IX.1 Sites internet</u>	111
<u>IX.2 Livres</u>	112
<u>IX.3 Autres support</u>	112
<u>X Annexe B : Glossaire</u>	113
<u>XI Annexe C: les données</u>	117
<u>XII Annexe D: Tests unitaires</u>	120
<u>XII.1 La maquette</u>	120
<u>XII.2 Tests unitaires de l'identification</u>	121
<u>XII.3 Gestion Molécule</u>	125
<u>XIII Annexe E : Organigramme hiérarchique</u>	129

Index des figures

Figure 1: Schéma simplifié de l'organisation de l'université.....	13
Figure 2: Réseau de l'université à Internet [PW3].....	17
Figure 3 : Réseau Schématisé de CPE	18
Figure 4: Arbre de décomposition d'une molécule.....	26
Figure 5 : Schéma conceptuel initial de données.....	32
Figure 6 : Schéma conceptuel de données après les premières optimisations.....	37
Figure 7 : Modèle physique de données.....	38
Figure 8: la représentation chimique de la molécule LIMA.....	41
Figure 9: Cycle de vie en V [WP15].....	63
Figure 10: Diagramme de GANTT du projet.....	65
Figure 11: Modèle Physique de données actuel.....	76
Figure 12: Schéma d'appel de la procédure récursive.....	86
Figure 13: Grandes étapes du traitement récursif.....	86
Figure 14: Schéma du plan du site CPSM.....	88
Figure 15: Bandeau de page.....	89
Figure 16: Page d'accueil initiale.....	89
Figure 17: Zoom de la page d'identification d'une personne.....	90
Figure 18: Zoom de la page d'identification : mot de passe perdu.....	91
Figure 19: Page d'accueil avec l'utilisateur identifié.....	92
Figure 20: Zoom de la page Molécules disponibles : onglets et listes.....	93
Figure 21: Vue globale de la page gestion d'une molécule.....	95
Figure 22: Liste des sous-molécules (zone 3).....	95
Figure 23: Zoom de la page : Liste de sous molécules disponibles.....	96
Figure 24: Fenêtre de modification d'un paramètre de requête.....	101
Figure 25: Bug d'affichage de Dreamweaver.....	104
Figure 26: Organigramme hiérarchique au niveaux des laboratoires.....	129

Index des tables

Tableau 1: Volumes molaire partiels et contributions aux paramètres de solubilité partiels des différents groupes chimiques.....	41
Tableau 2: Paramètre de solubilité de quelques solvants.....	42
Tableau 3: Extrait du tableau des contributions des groupes de composants aux paramètres de solubilité (Van Krevelen).....	44
Tableau 4: Extrait du tableau des valeurs recommandées pour le volume molaire.....	45
Tableau 5: Phase initiale du projet.....	62
Tableau 6: Phases réalisées du projet.....	64
Tableau 7: Plan tests unitaires de la maquette.....	99
Tableau 8: Plan des tests unitaires pour l'identification.....	100
Tableau 9: Plan des tests unitaires pour la gestion d'une molécule.....	100
Tableau 10: Contribution des groupes de composants aux paramètres de solubilité (Van Krevelen).....	118
Tableau 11: Valeurs recommandées pour les volumes molaires à 298 K.....	119
Tableau 12: Test de saisie du nombre de fragments.....	120
Tableau 13: Test sur l'affichage du calcul.....	121
Tableau 14: Test avec un exemple COO(CH ₂) ₅	121
Tableau 15: Test pour identification d'un utilisateur.....	122
Tableau 16: Test de changement d'utilisateur.....	122
Tableau 17: Test sur mauvais mot de passe.....	123
Tableau 18: Test de la création d'un utilisateur.....	123
Tableau 19: Test sur l'envoi du mot de passe.....	124
Tableau 20: Test de changement du mot de passe.....	124
Tableau 21: Tests unitaires de création d'une nouvelle molécule.....	126
Tableau 22: Test unitaires de sauvegarde d'une molécule.....	126
Tableau 23: Tests unitaires pour la sauvegarde des sous-molécules.....	127
Tableau 24: Tests unitaires pour la modification de sous-molécules.....	127

Projet C.P.S.M.

Les remerciements

Je tiens particulièrement à remercier mes encadrants qui, grâce à leur savoir et leur soutien m'ont permis de progresser et de présenter ce mémoire.

Merci au directeur de recherche Hamid Elaissari et à son équipe pour ce projet passionnant pour eux et pour moi, leur disponibilité et leur patience.

Merci à Hatem Fessi, directeur du LAGEP et à mes collègues, pour leurs encouragements.

A mes amis, toujours présents dans les moments de découragement.

Avec une mention spéciale à

Olivier pour sa ténacité et son optimisme.

Véronique pour sa patience et son courage.

Dominique pour son soutien et sa persévérance.

Michelle pour sa gentillesse et sa patience.

A la mémoire de mes parents, à qui je dédie ce travail.

Projet C.P.S.M.

I Introduction

Le laboratoire d'automatique et du génie des procédés, le LAGEP, est un laboratoire qui associe plusieurs disciplines dont principalement le génie des procédés et la pharmacotechnie. Il utilise de nombreux outils dans différents domaines comme l'automatique, la cristallisation ou la chimie. Un de leurs axes de recherche concerne la caractérisation des molécules afin de connaître leurs comportements, leurs conditions d'utilisation et leurs possibilités d'application. Un de ces paramètres est la solubilité dans un milieu.

La *solubilité* d'une substance chimique est la quantité maximale de cette substance que l'on peut dissoudre dans un solvant [PW1]. Afin de déterminer la solubilité d'un produit, il est nécessaire de faire de nombreuses expériences ou d'étudier les forces de la physico-chimie en action : des méthodes plus ou moins compliquées à mettre en œuvre et complexes à réaliser.

La véritable réponse à ce problème est l'expérience de chimie. Le chimiste doit diviser le produit étudié et tester méthodiquement ce produit dans un milieu cible, que l'on désigne sous le nom de solvant. Nous pouvons tout de suite imaginer la limite de cette méthode : le premier produit peut être difficile à obtenir et passer en revue l'ensemble des solvants augmente de manière importante le temps d'étude et la consommation de produit actif. Par conséquent, cette solution ne peut être retenue pour des raisons évidentes de temps et d'argent.

La seconde possibilité : l'étude des forces physico-chimique se complexifie très rapidement. La saisie de molécule de plus en plus longue nécessite des calculs des interactions entre atomes de plus en plus complexes et cette lourdeur de modélisation limite son utilisation.

A partir de cette deuxième possibilité, le problème peut être envisagé sous un angle différent : prendre les acquis des expériences et chercher une ou plusieurs méthodes pour pouvoir les combiner afin de produire un résultat, peut être plus approximatif, mais plus facile d'usage tant sur le point de la modélisation que celui de sa manipulation.

Cette solution innovante proposée par le directeur de recherche Hamid Elaissari n'a jamais été implémentée jusque là. Ce mémoire décrit le travail réalisé dans ce but, les moyens mis en œuvre pour obtenir une solution satisfaisante. Nous avons appelé ce projet « **Calcul du Paramètre de Solubilité Moléculaire** » (C.P.S.M.).

Je m'attacherai dans une première partie à décrire le contexte du projet avant d'effectuer un état des lieux. L'analyse du problème et la proposition de solutions fera l'objet de ma troisième partie. Une quatrième partie mettra en lumière le développement de la solution retenue. Une cinquième partie traitera de la mise en œuvre du projet. Les contraintes seront exposées en sixième partie. Enfin, je terminerai en faisant un bilan de cette étude.

II Contexte

Un projet de cette envergure ne peut être compris dans sa totalité si je ne donne pas le contexte dans lequel il s'insère. J'ai souhaité travailler dans mon milieu professionnel pour plusieurs raisons ; notamment pour avoir une plus grande implication dans la recherche liée à mon laboratoire. Je souhaitais aussi montrer que l'informatique, à travers mon savoir faire, peut apporter autre chose que les outils classiques utilisés par mon unité. Cette partie de document s'attache à décrire mon environnement de travail, et ceci, à différentes échelles du laboratoire à l'université.

II.1 Mon travail

Actuellement, j'occupe un poste d'ingénieur d'étude en informatique en temps partagé sur deux laboratoires qui ont des thèmes de recherche différents.

J'exerce mes fonctions à vingt pour cent, dans le laboratoire de génie électrique de Lyon, portant le nom d'AMPERE. Ce laboratoire est présent sur 3 établissements : l'École Centrale de Lyon (ECL), l'Institut National des Sciences Appliquées (INSA), et l'Université Claude Bernard Lyon 1 (UCBL1). Sur ce dernier site, j'ai la responsabilité de maintenir leur système d'information, de trouver des solutions et résoudre tous leurs problèmes informatiques : du plus insignifiant au plus complexe, tant sur du ponctuel que sur des projets à long termes. Je travaille en collaboration avec un autre informaticien, lui aussi à vingt pour cent sur ce site. Quatre professeurs, neuf maîtres de conférence, un chargé de recherche, deux BIATOSS et neuf étudiants-chercheurs sont actuellement sur le site.

Je consacre quatre-vingts pour cent de mon temps de travail dans le laboratoire d'automatique et de génie des procédés, le LAGEP, dont le cœur de métier est l'interaction entre différents domaines que sont le génie des procédés, la pharmacotechnie et l'automatique. Le génie des procédés touche les domaines de la cristallisation, du séchage et du changement d'échelle, c'est à dire, du transfert du laboratoire vers l'industrie. En tant qu'Unité Mixte de Recherche (UMR), nous accueillons aussi bien des enseignant-chercheurs, que des chercheurs du Centre National de la Recherche Scientifique (CNRS).

Ce laboratoire compte une centaine de personnes dont je suis le seul informaticien. Par conséquent, j'assume de multiples responsabilités notamment celui d'administrateur système et réseaux. J'ai la responsabilité de maintenir fonctionnelle une partie du réseau Ethernet. Je gère une plage d'adresses IP, sachant que je n'ai pas l'autorisation de la part de l'école de CPE, pour l'instant, d'avoir un serveur DHCP qui me permettrait d'affecter plus efficacement et automatiquement les configurations réseaux. J'assure le bon fonctionnement des différents serveurs de l'unité comprenant des serveurs web, le serveur mail, le serveur de partages de fichiers et le serveur du domaine Windows. Je gère les comptes nécessaires à l'accès et l'utilisation de ces ressources serveurs. J'assume aussi la gestion des systèmes d'impressions et de sauvegarde. Je m'occupe, également, du parc informatique. Je conseille les enseignants chercheurs des deux laboratoires en prenant en compte à la fois leurs besoins, la destination première du matériel (machine de calcul ou d'expérimentation), le respect des contraintes qui nous sont imposées par les établissements comme les marchés informatiques ou les sources des financements. Les machines personnelles sont installées en fonction des demandes spécifiques des chercheurs. Les utilisateurs se divisent en trois grandes catégories : les automaticiens, gros consommateurs de logiciels comme Matlab et ses toolbox ; ceux du génie du procédé, utilisant des logiciels orientés vers la chimie et du développement Fortran ou Labview ; et les pharmaciens, gros consommateurs de logiciels bureautiques. Le parc informatique est de ce fait particulièrement hétérogène, le personnel ne se sent pas concerné par les contraintes liées à l'informatique. Afin de me permettre la réalisation de ma mission, le directeur m'a alloué un budget spécifique annuel que je gère à ma guise.

Interlocuteur privilégié, je concentre toutes les demandes d'informatique et de sécurité du système d'information. J'ai donc de nombreux contacts avec les fournisseurs, pour le suivi du matériel, des pièces et des réparations. Je suis sollicité dès que cela touche de près ou de loin le domaine informatique, et parfois à tort. Grâce aux listes de diffusion, aux colloques et aux formations, je suis dans les réseaux d'information de mon domaine des différents établissements auxquels mon unité appartient.

En parallèle, après quelques années, j'ai souhaité pouvoir participer à la recherche et m'investir dans des projets de développement. J'ai donc réussi à dégager une partie non négligeable de mon temps, avec l'accord de mon directeur de laboratoire principal, le

professeur Hatem FESSI. Ceci me permet de réaliser des actions à plus long terme, d'apporter mon savoir faire informatique aux équipes des laboratoires et de cette manière de participer à leurs recherches et aux publications.

II.2 Rôle dans l'université

L'université a de nombreuses composantes. Afin de mieux comprendre mes activités, je vous propose un schéma simplifié de l'organisation de l'UCBL1. La description débute par l'établissement, en haut et finit par les laboratoires où je suis, en bleu et en bas.

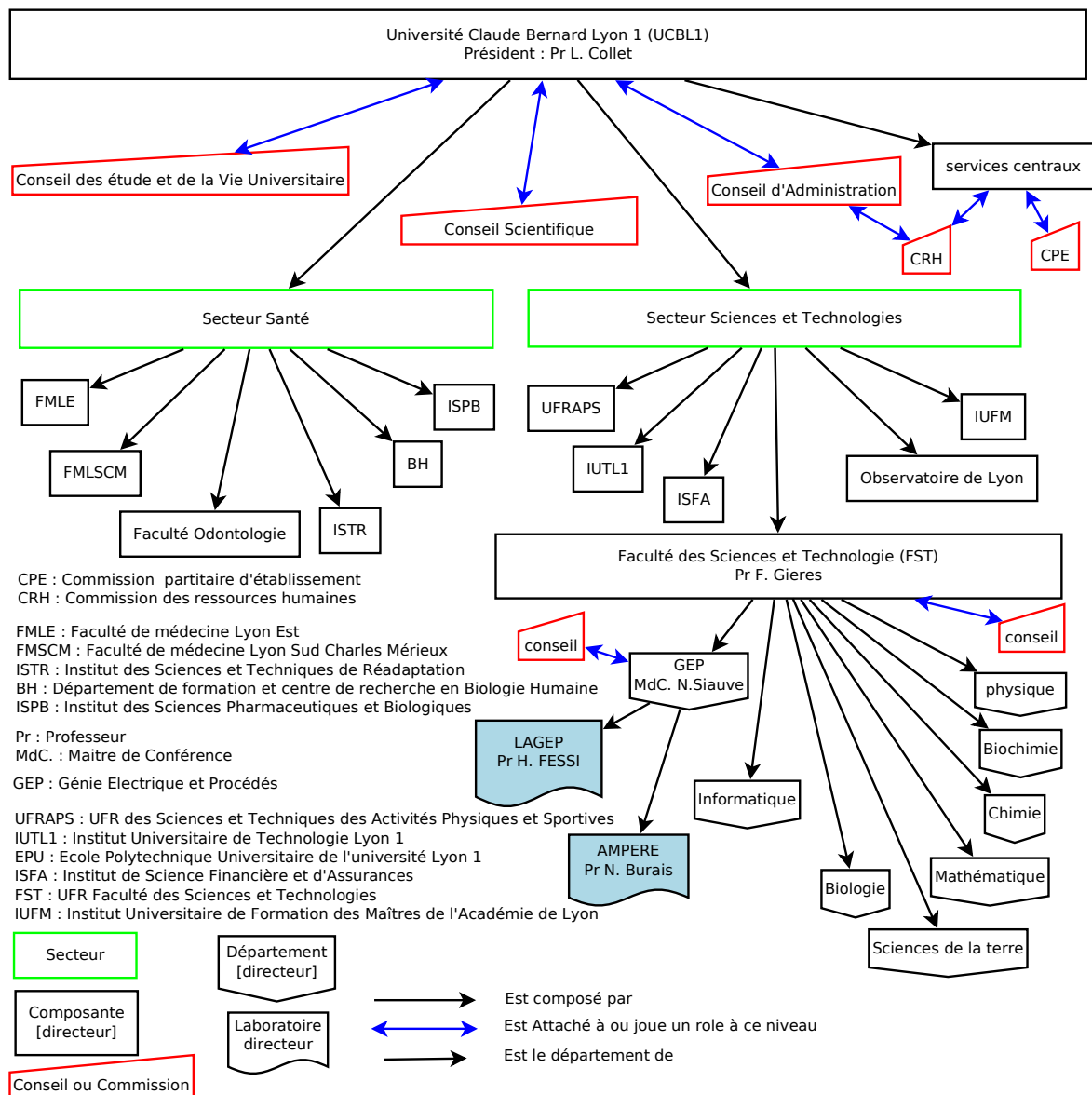


Figure 1: Schéma simplifié de l'organisation de l'université

J'occupe un poste à responsabilité sur deux laboratoires et ceux-ci inclus dans un regroupement de laboratoires, une ex-UFR devenue département (cf XIII. Annexe E : Organigramme hiérarchique page 129). J'ai tout naturellement eu à intervenir et à m'investir dans les structures de l'université. Représentant à tous les niveaux organisationnels de l'université, je considère comme faisant partie de mon travail de défendre les intérêts des personnes qui m'ont élu ou se sont fédérés derrière moi, ou les instances dans lesquelles j'ai un rôle comme celle de l'UFR « Faculté des Sciences et Technologies » (FST). C'est un travail d'importance qui demande du temps et qui doit être effectué de manière continue.

En effet, l'information circule difficilement entre les différentes strates de l'établissement, et perdre le fil des discussions peut faire rater des opportunités et provoquer des incompréhensions au sujet de certaines actions qui semblent tomber parfois du ciel. Je suis actuellement élu à la nouvelle UFR FST, regroupant l'ensemble des huit anciennes UFR de sciences ; toutes transformées en département. Je participe à différentes commissions et au conseil de cette « faculté ». Je suis aussi élu à la commission d'établissement paritaire (CPE aussi, mais pas l'école), qui statue sur les cas individuels pour le personnel de soutien à la recherche dont je fais partie. Je suis représentant de la FST à la commission des ressources humaines (CRH) émanant du conseil d'administration de l'université pour réfléchir sur la politique de gestion du personnel au niveau de l'établissement.

Sur mon lieu de travail, j'ai eu le plaisir de rencontrer des élèves du Cnam qui ont effectué leurs stages de fin de cycle. Ce qui m'a décidé à entamer mon cursus Cnam et après quelques années d'effort, j'ai l'honneur de vous présenter mon travail sous forme de ce mémoire.

III État des lieux

Ce chapitre présente l'état des lieux au moment où le projet a débuté. Je donne une image des acteurs, des informations disponibles et de l'expression du besoin à ce moment. Je commencerai par présenter les acteurs et je finirai par décrire le système informatique.

III.1 Personnes et besoins

Ce projet démarre par un besoin exprimé par le directeur de recherche nouvellement arrivé au LAGEP, monsieur Hamid Elaissari. Celui-ci anime une équipe de douze chercheurs, travaillant dans le domaine de la chimie à la frontière du génie pharmaceutique et du génie des procédés. Ces recherches l'ont poussé à travailler sur le domaine de la solubilité. Il s'est aperçu qu'il n'existe pas de solution simple pour calculer le paramètre de solubilité d'un produit. Il m'a donc proposé d'étudier la réalisation d'un logiciel évolutif répondant à cette problématique.

Les travaux sur les calculs de solubilité intéressent particulièrement Claudia Elizabeth Mora Huertas pour ses travaux de recherche et sa thèse [MH11]. Elle étudie la création de nanocapsules [PW2]. Les nanocapsules sont de petites capsules, de l'ordre des nanomètres, qui contiennent un produit et sont constituées d'une membrane polymère. Elles sont fabriquées d'une façon précise et elles encapsulent un produit actif à usage thérapeutique, comme un médicament ou des biomolécules. Elles peuvent être déposées ou transportées vers une cible, un organe. L'objectif est qu'elles se dissolvent progressivement et libèrent le principe actif de façon continu. Pour produire les nanocapsules, il faut étudier les solubilités de leurs membranes et du produit actif afin de trouver les bons solvants dans lesquels on peut les dissoudre totalement sans formation de cristaux. Ceux-ci doivent être utilisables dans le cadre de produit à destination du vivant et compatibles avec les normes pharmaceutiques. Au début du projet, Mlle Mora Huertas calculait l'ensemble de ses valeurs à la main. D'autres doctorants vont avoir rapidement besoin de réaliser le même genre de calcul d'où l'intérêt de ce projet.

III.2 Système informatique

Le système informatique du LAGEP est logiquement structuré autour de deux infrastructures principales qui sont les hébergeurs géographiques actuels du laboratoire. L'école supérieure CPE (Chimie Physique Électronique), une école d'ingénieurs qui se situe sur le campus de la Doua, et l'université Claude Bernard, dans laquelle se situe l'école.

Le LAGEP possède trois serveurs opérationnels. Ces machines sont dans le domaine de nom de l'école. Le nom de domaine est « cpe.fr », lui même hébergé par l'université Lyon 1, dont le nom de domaine est « univ-lyon1.fr ». Chaque domaine possède son propre pare-feu et ses règles de filtrages.

Afin de vous donner une vision d'ensemble, je décris dans la figure 2 ci-après l'organisation de l'infrastructure informatique de l'université, puis celle de CPE avec une seconde figure. Tous ces réseaux contribuent à permettre au LAGEP de pouvoir accéder à internet et d'être accessible depuis celui-ci.

L'université est reliée à internet via le **Réseau Optique du Campus de la Doua** (ROCAD), puis le Lyres (**Lyon Recherche Enseignement Supérieur**, le réseau métropolitain de l'agglomération lyonnaise), et enfin, le **Réseau National pour la Technologie, l'Enseignement supérieur et la Recherche** (Renater).

ROCAD est la liaison entre l'université Lyon 1 et l'école de CPE et est géré par le CISR (**Centre Inter-établissement pour les Service Réseaux**). Ce réseau pourvoit aux communications entre RMU et les différents établissements dont l'école CPE. Le LAGEP est fortement dépendant des choix d'infrastructure de CPE, étant hébergé en son sein.

Aperçu du réseau : de l'université Claude Bernard à Internet

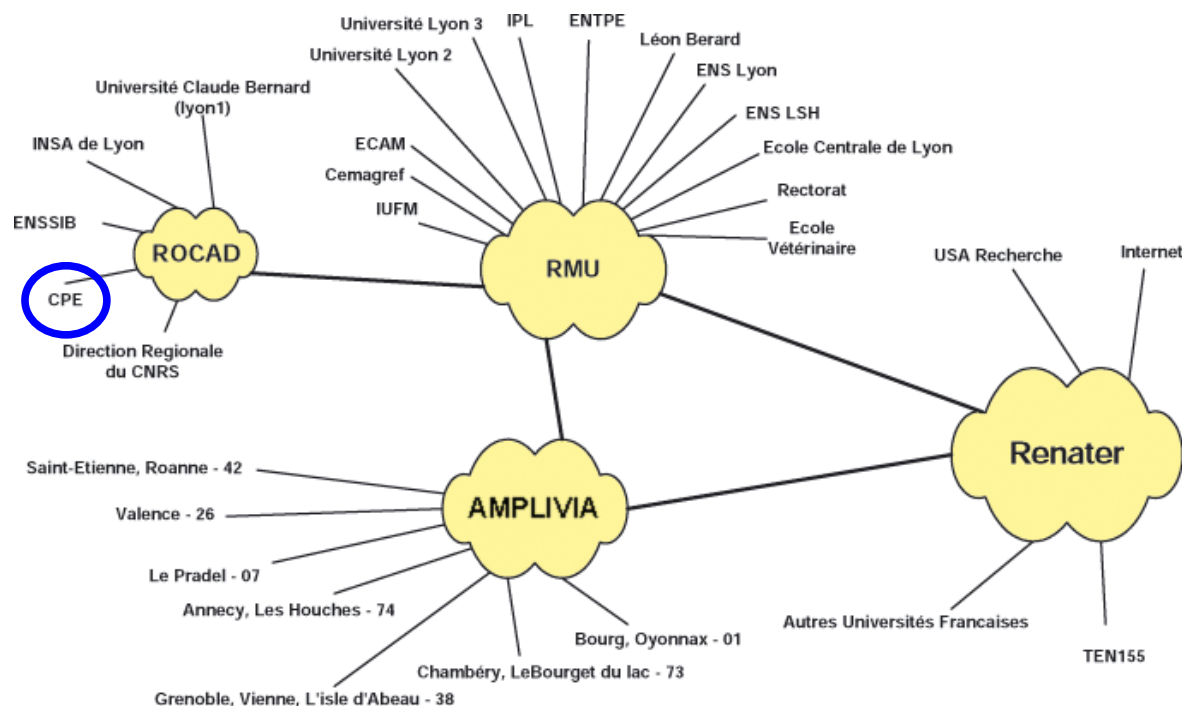


Figure 2: Réseau de l'université à Internet [PW3]

ROCAD : Réseau Optique du CAmpus de la Doua

RMU : Réseau Métropolitain Universitaire

AMPLIVIA : réseau régional haut-débits

Renater : Réseau National pour la technologie, l'enseignement supérieur et la recherche

Actuellement, le LAGEP est client du réseau de CPE. Celui-ci est représenté par le schéma figurant à la page suivante. Occupant plusieurs étages, nous partageons les différentes sections Ethernet des bâtiments. La réception se fait par une liaison ATM (Asynchronous Transfer Mode) depuis ROCAD.

Après le pare-feu, un routeur distribue les connexions réseaux sur des appareils actifs situés à chaque étage. Ceux-ci, avec des commutateurs empilés, font la répartition vers les différentes prises de l'étage.

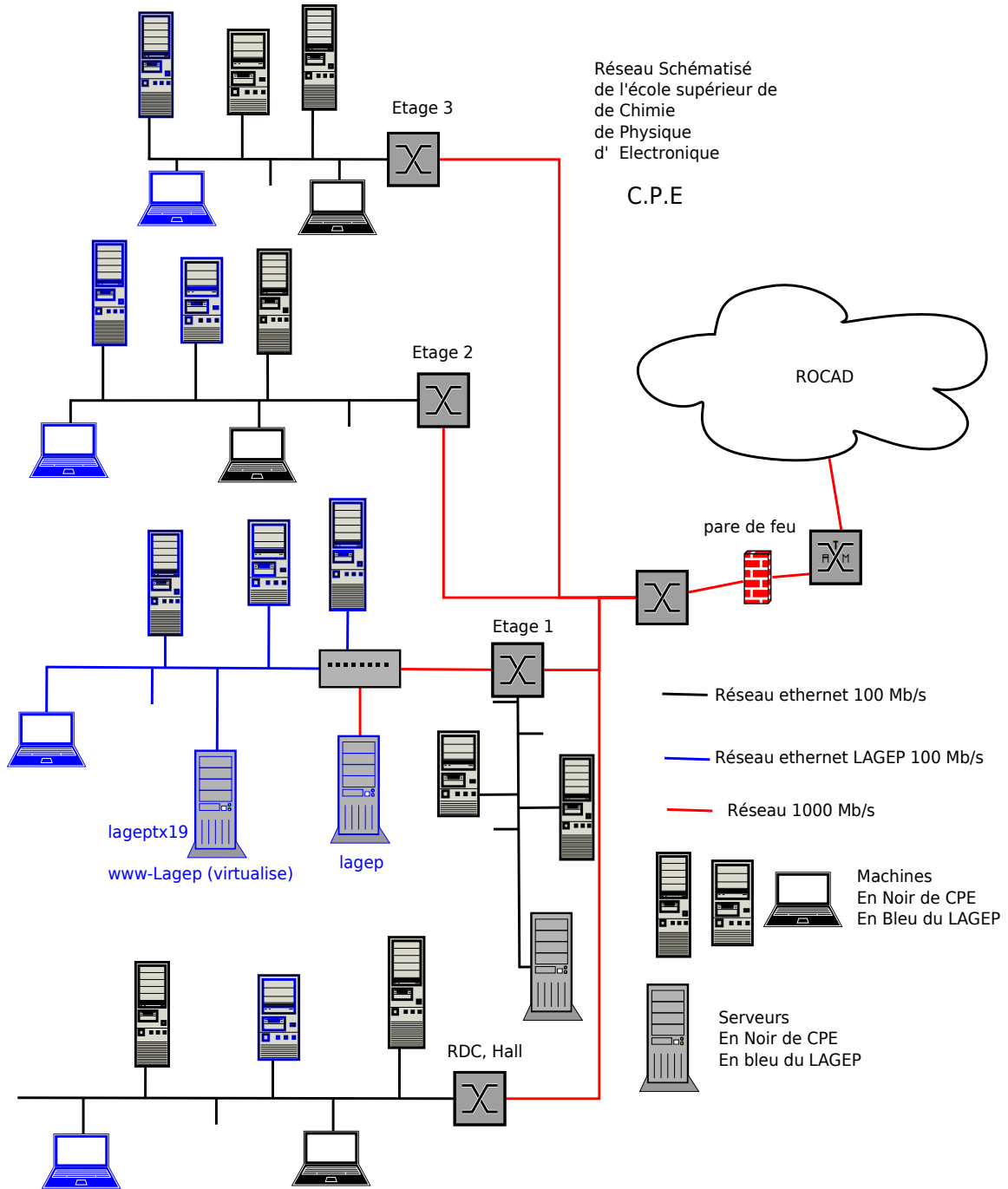


Figure 3 : Réseau Schématisé de CPE

Au niveau du pare-feu de CPE et de l'université Claude Bernard, des règles ont été écrites pour que, depuis l'extérieur et pour le LAGEP, seul le serveur web soit accessible sur les protocoles http et https. Il possède le nom de *www-lagep.cpe.fr*, avec différents alias dont *www.lagep.univ-lyon1.fr*. Ce n'est pas le seul serveur visible de l'extérieur, mais c'est le seul qui soit visible comme serveur web sans périmètre restreint.

Ainsi, un deuxième serveur, aux règles plus complexes, est aussi visible de l'extérieur. Il permet l'accès aux différents protocoles pour les relevés de courriels. Mais bien qu'il héberge un serveur web, celui-ci est destiné aux tests et aux publications de sites web de type « intranet ». Sa visibilité a été réduite pour n'être visible que sur le réseau « cpe.fr » et « univ-lyon1.fr ». Afin de faciliter les interventions sur tous les serveurs du LAGEP, le service sécurisé SSH est autorisé sur les deux machines pré-citées depuis n'importe quelle adresse IP.

Suite à de nouvelles activités de recherche au niveau de CPE, et aux départs de certains laboratoires qui disposeront de nouveaux bâtiments, une réorganisation interne des laboratoires a été proposée. Le LAGEP étant le plus gros laboratoire de l'école supérieure de Chimie Physique et Électronique (ESCPE ou CPE), avec de plus en plus de chercheurs accueillis en son sein, il a été réparti sur tous les étages. Pour ce faire, un changement d'étage de la salle serveur et de la salle informatique, incluant des bureaux et un atelier, a dû être planifié et réalisé. Le matériel informatique a été déplacé du troisième étage au premier. Ces changements ont fait apparaître des bénéfices supplémentaires.

Jusqu'au mois de septembre, mon statut de client du réseau de CPE me permettait aucun contrôle direct sur le réseau. Depuis l'installation dans les locaux rénovés pour l'occasion, j'ai la possibilité d'activer ou de modifier cette partie du système d'information. Bien que nous soyons installés maintenant sur quatre étages, le premier est sous mon contrôle. J'ai la responsabilité de l'allocation des prises téléphoniques et connexions réseaux.

Mais surtout, j'ai accès à la baie réseau, cela me permettra d'envisager, grâce à un commutateur manageable acheté par le laboratoire, d'améliorer et de contrôler cette section du réseau : première pierre de la construction du réseau propre au LAGEP.

Enfin, j'ai eu un changement de débit réseau, ce qui me permet d'avoir une liaison en gigabit. Avec l'appareil actif fourni par CPE, je peux continuer cette liaison jusqu'au serveur principal de mon laboratoire.

Un projet à plus long terme, tant politique qu'informatique, prévoit un réseau uniquement LAGEP, à travers des éléments actifs nous appartenant. Avec un seul point

d'entrée entre nous et CPE, nous n'aurions plus à partager le réseau avec d'autres postes et d'autres laboratoires. Mais les difficultés sont nombreuses, car les capacités financières entre le LAGEP, l'université et CPE sont très différentes. De plus, les marchés, la gamme de produits, les prises de responsabilités suivent des circuits différents.

Pour toute intervention sur la topologie à l'extérieur de cette section, ou sur le changement des règles de filtrage, je dois faire une demande à l'équipe informatique de l'école. Étant aussi le correspondant sécurité du système d'information (CSSI) du laboratoire, pour l'université et pour l'école, je dois passer obligatoirement et logiquement par le responsable de l'informatique de CPE pour toute modification des règles de filtrage afin que celui-ci les valide et les transmette à l'université.

Enfin, le déplacement de la salle serveur m'a permis de rationaliser un peu plus l'organisation du système informatique. Dans les faits, cela s'exprime principalement par la virtualisation du serveur web.

Celui-ci était installé sur machine avec une architecture vieillissante et bruyante. Il était devenu obsolète aussi bien par la puissance de traitement, que par le manque d'espace disque. Habituellement, j'utilisais l'unité centrale la moins puissante pour installer ce type de serveur, car celui-ci n'est que très peu sollicité. Un serveur accessible au monde entier sur un protocole aussi courant devait être isolé. Je souhaite que ce serveur n'ai que le minimum de service pour faciliter toute réparation ou intervention en cas de panne ou d'attaque. Pour améliorer celui-ci, j'avais la possibilité de le migrer vers un serveur plus récent rendu disponible. Cependant, j'ai préféré opter pour la virtualisation, pour les deux raisons suivantes. Je voulais tester les produits de virtualisation en tant que serveur, surtout ceux proposés en Open Source et, me rendre indépendant du choix matériel. Car ce serveur WEB, au niveau du laboratoire, sert principalement à présenter celui-ci à des fins de vitrine. Il héberge peu de sites Web et nécessite un processeur peu puissant, un espace disque de bonne dimension et une carte réseau pour communiquer ses réponses aux requêtes des clients distants. Donc, il ne dépend pas réellement d'une infrastructure matériel. En cas de problème, je peux installer ce serveur élémentaire sur d'autres machines.

En résumé, le LAGEP a actuellement les serveurs suivants :

- le serveur Lagep cumule le rôle de serveur de comptes, de serveur de fichiers, de domaine SAMBA, de messagerie, d'impression, serveur web "intranet" tel que je l'ai décrit ci-avant. Il est sous debian lenny 5.0rc. Le système est sur des disques en raid 1, les données sur deux raids 5, avec des volumes gérés en LVM.
- le serveur, de nom lageptx19, est le serveur de base de données et d'application tierce. A l'origine, ce serveur devait permettre de lancer des calculs ou des programmes au temps de calcul long et lourd. La recherche du LAGEP n'a plus besoin de ce genre de calculs centralisés, les machines personnelles font très bien l'affaire pour leurs modélisations et leurs programmes. J'ai donc implanté un serveur de base de données, exploitant actuellement dix bases à l'aide du moteur de base de données MySQL version 5. Ayant encore beaucoup de ressources, après un temps de test, j'ai ajouté un service de virtualisation. C'est un serveur linux, debian lenny 5.0rc, disque en miroir (RAID 1)
- le serveur www-lagep est le troisième serveur, son adresse étant visible sur Internet, je l'estime plus sensible aux attaques extérieures que les autres machines. Il est limité à un rôle de serveur web. Il a été virtualisé cet automne sur le serveur Lageptx19, comme je l'ai expliqué au paragraphe précédent. C'est une debian 5.0c, sur noyau linux.

IV Analyse et solutions envisageables

Dans un premier temps, ce chapitre explique le besoin exprimé par les chercheurs, l'analyse que j'ai pu en déduire avec leur aide. Ensuite, je proposerai une analyse sur les données à traiter. La partie suivante présente une explication de la solubilité et les différents calculs et paramètres, en finissant par les calculs des paramètres de solubilités. Enfin, je compléterai l'analyse en deux points, l'analyse des traitements et la reprise du cahier des charges.

IV.1 Expression des besoins

L'objectif est de créer un moyen de faciliter le calcul du paramètre de solubilité d'une molécule, plus exactement, d'une structure moléculaire, afin de connaître l'affinité avec un milieu, désigné par le mot "solvant" [PW1]. L'idée est de permettre le calcul sur des structures de plus en plus grandes en laissant les choix aux utilisateurs, chimistes ou associés, des parties de molécules qu'ils souhaitent étudier.

Actuellement, il n'existe pas de logiciel réalisant ces calculs, dans les conditions d'entrées proposées c'est à dire à partir d'une description de la molécule à l'aide de fragments de molécule plus petites. Des publications parlent souvent de logiciel développé maison en Matlab ou en Fortran ; généralement dédié à une catégorie de produits chimiques, comme le charbon ou des logiciels à destination d'une application spécifique comme la peinture, la dispersion de polymère. M. Elaissari souhaiterait, donc, une application disponible, avec une interface facile à utiliser, donnant le maximum d'informations pour aider à vérifier le résultat ou permettre de comprendre l'importance ou le poids de certains résultats vis à vis du résultat final. Le directeur de recherche souhaite aussi que l'on puisse valider les informations entrées à posteriori, c'est à dire les valeurs proposées par l'utilisateur, afin d'augmenter le référentiel. L'utilisateur pourrait proposer une molécule, ses valeurs d'expérimentation et de donner les sources lui permettant de valider ses résultats. Car il existe deux sources à ces valeurs, la première est l'expérimentation personnelle, la seconde est la bibliographie.

IV.2 Analyse des besoins

Je propose d'étudier l'analyse des besoins. Afin de réaliser cet objectif, je résumerai les besoins, puis mettrai en avant chaque élément de phrase de ce résumé pour donner des explications et faire des propositions de fonctionnalités.

Je peux résumer les besoins de la manière suivante :

A partir de plusieurs sources, permettre le calcul du paramètre de solubilité, en suivant plusieurs méthodes et ceci au niveau moléculaire. L'accès aux calculs, à la saisie et à l'exploitation des molécules doit être simple, tout en respectant la « propriété » et les modifications de chacun. Un référentiel sera fourni.

Nous pouvons détailler ce résumé, en découpant le texte ci-dessus pour définir chaque groupe grammatical, un paragraphe ayant pour titre l'extrait de la phrase.

IV.2.1 A partir de plusieurs sources

Il est important de donner l'origine, c'est à dire la source des valeurs fournies, afin de permettre de connaître la pertinence ou la fiabilité d'un calcul. Connaître les sources de ces valeurs donnera la meilleure garantie scientifique possible. Mais, le chimiste ou l'utilisateur final de l'application pourra saisir ses propres résultats afin de les utiliser pour ses calculs. Il aura la possibilité d'exploiter ses résultats dans l'estimation des paramètres de solubilité de molécules plus « grosses ». Le scientifique devra donc bien percevoir l'origine des valeurs de ses molécules vis à vis de la construction moléculaire qu'il a choisie ou déterminée.

IV.2.2 Plusieurs méthodes

Bien que la méthode de Van Krevelen (développée plus loin) soit prise comme référence, il existe d'autres méthodes, par exemple, celle de Small et celle de Hoy. Il faudra prendre en compte la possibilité d'ajouter ces méthodes dans le logiciel. Donc, il sera nécessaire de savoir quelles sont les valeurs utiles aux calculs. Le logiciel devra prévoir de ne pas stocker plusieurs fois les mêmes valeurs utilisées qui peuvent apparaître dans plusieurs calculs. Il faudra déterminer les règles de calculs et des transmissions de valeurs liées à la

méthode de décomposition de molécules (voir plus bas, Les molécules).

IV.2.3 L'accès aux calculs

Le logiciel proposera une présentation des calculs. L'objectif est d'expliquer les résultats, en montrant les valeurs intermédiaires, et connaître le poids de ces valeurs dans le résultat final. Les utilisateurs pourront donc vérifier leurs valeurs et la source de ces valeurs. Si plusieurs méthodes sont proposées, chacune aura ses valeurs affichées dans sa partie. Dans le cas où le logiciel ne pourra donner de valeur pour une méthode car certaines valeurs ne sont pas connues au moment du calcul, l'application indiquera clairement les blocages, par exemple, quelles sous-molécules ne possèdent pas les valeurs requises.

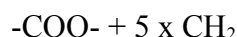
IV.2.4 Les molécules

Une molécule est une séquence cohérente d'atomes qui suit les règles de la chimie et de la physico-chimie. Certaines molécules ont des structures répétitives, d'autres des longueurs importantes, mais toutes sont composées d'ensembles cohérents de séquences d'atomes, donc d'un ensemble de molécules : les sous-molécules.

D'un point de vue d'un graphe, une molécule est un arbre de décomposition, dont les feuilles doivent être connues et les arcs relient une molécule à l'ensemble des sous-molécules. Chaque nœud de cet arbre est une sous-molécule. En tant qu'ensemble, la décomposition ne prend pas en compte l'ordre ou la séquence d'apparition des sous-molécules dans la description de la molécule, seul compte le nombre d'occurrences. Chaque arc est pondéré par le nombre de sous-molécules.

Exemple : le **Polycaprolactone**, possède une structure moléculaire de la forme $\text{COO}(\text{CH}_2)_5$ que l'on note $[\text{COO}(\text{CH}_2)_5]$.

En supposant que dans le référentiel choisi, les valeurs pour -COO- et pour CH_2 sont connues. L'utilisateur pourra décomposer sa molécule en deux feuilles :



ce qui correspond à l'arbre suivant :

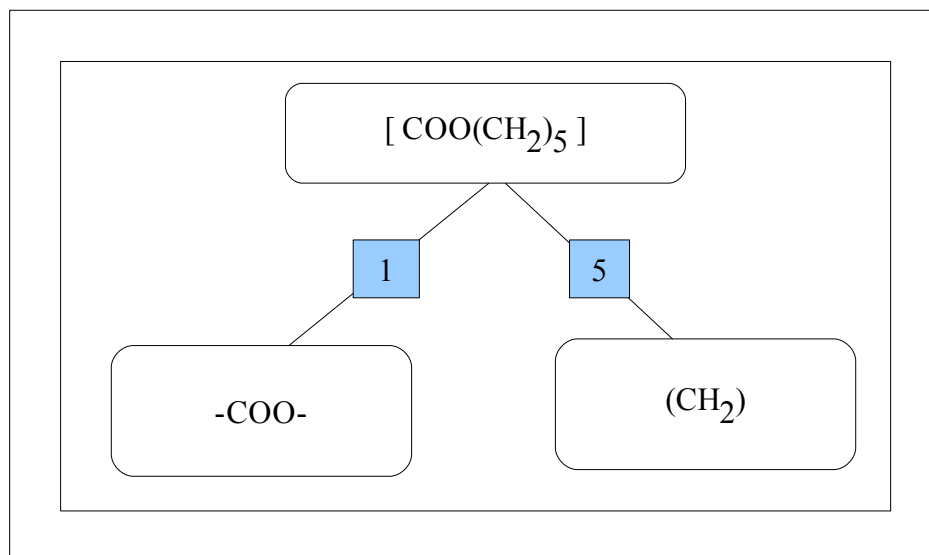


Figure 4: Arbre de décomposition d'une molécule

Pour information, -COO- est une sous-molécule de $[\text{COO}(\text{CH}_2)_5]$.

IV.2.5 La saisie et l'exploitation

La gestion des molécules devra être simple, mais pas simpliste. L'utilisateur saisira ses molécules et pourra, dans certains cas, rappeler celles-ci afin de connaître leurs compositions. Il faut donc, au minimum, que l'application permette la saisie, la sauvegarde, le rappel et la modification d'une molécule. Il faudra faire bien attention à la suppression de molécule, pour ne pas supprimer des molécules qui sont sous-molécules d'autres molécules. Il faudra aussi prendre soin d'éviter toute possibilité de cycle ; une molécule devient, par des saisies successives et par transitivité, sa propre sous-molécule.

IV.2.6 La propriété

La notion de propriété est une notion importante. Les chercheurs ne souhaitent pas

toujours partager leurs méthodes ou leurs choix de décompositions. Ils peuvent aussi préférer utiliser leurs propres valeurs, résultat de leur recherche pour une molécule particulière. La propriété donne le droit de modifier la molécule, et par extension, l'ensemble des molécules qui référencent celle-ci.

De plus, elle est intrinsèquement liée à la notion de visibilité. C'est à dire que la visibilité d'une molécule dépend des groupes. Par défaut, seul le propriétaire peut voir la molécule, mais il peut souhaiter partager celle-ci avec d'autres personnes. Dès lors, le mécanisme de visibilité devient complexe.

Chaque utilisateur peut appartenir à un ou plusieurs groupes. Le propriétaire de la molécule peut rendre visible pour un ou plusieurs groupes une molécule dans le but de partager son travail avec les membres de son laboratoire, par exemple, afin de réaliser un référentiel commun avec ceux-ci. Il existe donc une relation entre une molécule et un utilisateur qui traduit la notion de visibilité. L'utilisateur peut voir toutes molécules affectées à un groupe auquel il appartient. Mais cela implique d'autres problématiques : que faire si une molécule devient invisible à un groupe et qu'elle est toujours utilisée comme sous molécule d'une des siennes ? Faut-il prévenir l'utilisateur ? Faut-il prévenir le propriétaire que sa molécule ne peut pas être rendue invisible et permettre la visualisation de cette molécule ? Autoriser à faire une copie ? Le logiciel aura clairement un comportement défini en fonction de ces situations. Ainsi, le propriétaire connaîtra à l'avance les risques du partage de ses molécules et les autres, celui d'utiliser cette molécule.

Par ailleurs, l'utilisation de groupe, permet dans une certaine mesure de gérer les droits un peu plus facilement que si le propriétaire devait autoriser chaque personne individuellement. Ainsi, toute nouvelle personne qui s'inscrit dans un groupe, pourra profiter des droits sur l'ensemble des molécules visibles de ce groupe sans qu'il soit nécessaire que tous les propriétaires donnent leur accord. Une procédure d'inscription à un groupe doit être réalisée et mise en place. Les administrateurs de groupe étant les seuls à pouvoir inscrire ou révoquer un utilisateur. Le créateur du groupe devient de facto le premier administrateur. Un administrateur n'est pas forcément dans le groupe en tant qu'utilisateur. Il peut se désinscrire volontairement. Ainsi, certains comptes pourront, au besoin des utilisateurs, être spécialisés

dans la gestion des droits et de l'administration de groupes.

IV.2.7 Un référentiel sera fourni

Un ensemble particulier de molécules, correspondant à notre référence bibliographique, sera proposé dès l'installation de l'application. Ces molécules seront disponibles pour l'ensemble des utilisateurs, et seront généralement utilisées comme feuilles terminales de l'arbre de décomposition. Pour les distinguer des autres molécules, je les désigne sous le nom de fragments.

Cet ensemble pourra être enrichi par de nouvelles molécules, après validation et vérification des administrateurs scientifiques de l'application. Ce sont des comptes spécifiques et ils ne sont pas forcément administrateurs de groupes comme décrit dans le paragraphe précédent.

Par ailleurs, les fragments ne peuvent pas être modifiés par les autres utilisateurs, compte tenu de leur importance. Mais, si le besoin se fait sentir, un fragment pourra être copié afin d'être modifié et exploité. Ainsi, l'utilisateur pourra essayer de nouvelles hypothèses et continuer à exploiter le logiciel avec sa propre version du fragment.

Je viens d'exprimer et de détailler le besoin, mais cette simple analyse doit être complétée. Les parties suivantes du chapitre présenteront l'étude de l'existant, donnant un aperçu des différentes solutions utilisées habituellement pour connaître le paramètre de solubilité. Après, j'exposerai l'étude du besoin en vue d'obtenir les entités et les relations. Ceci me permettra de proposer un modèle conceptuel de données et un modèle physique des données. Je continuerai par donner les éléments nécessaires à la compréhension du calcul du paramètre de solubilité. J'exposerai la théorie du paramètre de solubilité, en reprenant l'explication donnée dans la thèse de [RJ94], avec un exemple. Enfin, je donnerai les différentes méthodes de calculs à envisager pour le logiciel.

IV.3 Étude de l'existant

Il existe actuellement des logiciels, comme « Molecular Modeling Pro »[PW4] qui permettent de calculer les énergies de covalence après modélisation, et vont jusqu'à proposer des valeurs pour le paramètre de solubilité. Ils nécessitent que l'utilisateur modélise précisément les différents atomes et leurs liaisons hydrogènes. La molécule est souvent modélisée en trois dimensions, avec une interface qui n'est pas toujours facile à utiliser. Comme je l'exposerai plus loin, la valeur du paramètre de solubilité est un calcul de plusieurs forces en action, s'influençant les unes des autres. A l'aide d'un modèle en trois dimensions, le logiciel peut donner des valeurs à chacune de ces forces et donc déduire le paramètre de solubilité, ainsi que d'autres valeurs ; le paramètre de solubilité n'étant pas leur objectif premier. Ces logiciels ne répondent pas au besoin exprimé parce que leurs complexités d'entrée, la modélisation en trois dimensions, sont très chronophages et peuvent être éloignées de l'expérimentation.

Par ailleurs, les calculs étant assez simples, ils utilisent des livres de référence pour la recherche de valeurs. Il faut donc que le chimiste parcourt ses livres afin de trouver la bonne décomposition, puis qu'il fasse les sommes des différents éléments de calculs (les forces en jeux, voir ci après). Le logiciel tentera d'automatiser ces opérations qui peuvent être fastidieuses et génératrices d'erreurs.

Certains chercheurs ont développé des logiciels à l'aide de langage, comme Fortran ou d'outils de programmation comme Matlab, de la société Matwork. Ils créent ces logiciels pour répondre à des besoins spécifiques, correspondants à des résultats dans des domaines précis, par exemple un modèle bi-phasique d'absorption ou le calcul, par estimation, du paramètre de solubilité de 1228 solvants [FG07] suivant une méthode algorithmique précise. Ces logiciels peuvent être utiles pour connaître leurs méthodes de saisies ou de calculs, mais sont généralement peu exploitables dès que nous sortons de leurs domaines d'applications.

Ainsi, l'exclusivité de l'application semble être corroborée par les recherches que nous, les chercheurs de l'équipe et moi-même, avons faites. Je vais donc présenter l'étude des besoins nécessaire pour la réalisation du logiciel.

IV.4 Étude des besoins

A partir de l'expression des besoins réalisée au sous chapitre IV.2, Analyse des besoins, j'ai fait ressortir les points suivants :

1. Un référentiel donnera des molécules élémentaires ou fragments. Celles-ci ont des valeurs connues et de sources connues. Elles contiennent l'information nécessaire aux calculs, elles sont donc les variables de ces calculs.
2. La capacité de créer de nouvelles molécules à partir des combinaisons d'autres molécules est demandée. L'ordre ou la séquence n'est pas important, mais le nombre de chaque sous-molécule doit être pris en compte.
3. La nouvelle molécule appartient à son créateur, qui peut la modifier ou la mettre à disposition d'autres utilisateurs.
4. Plusieurs méthodes de calculs peuvent être implémentées.
5. L'utilisateur pourra faire des copies, ou dupliquer des molécules et changer les valeurs calculées pour l'aider à réaliser des hypothèses, avoir ses propres exemplaires de molécules
6. Les molécules sont visibles pour certains groupes d'utilisateurs.
7. Un utilisateur peut appartenir à différents groupes, il devra en faire la demande aux administrateurs afin de s'insérer dans le groupe.
8. L'objectif de l'application est de calculer, de façon plus ou moins approximative, le paramètre de solubilité et de fournir une liste de solvants, proches de cette valeur, indiquant que la molécule a des affinités de solubilité avec ceux-ci.

D'après ce qui a été exposé ci-dessus, je peux exprimer le besoin de façon à mettre en avant les entités relations, à travers les phrases suivantes :

- Un fragment est une molécule du référentiel
- Une molécule est composée soit de fragments soit de molécules existantes.
- Une molécule possède des valeurs d'exploitation (désignées comme valeurs).
- Une valeur est associée à un ou plusieurs calculs.

- Chaque valeur possède une valeur numérique associée à une molécule avec une source bibliographique ou expérimentale.
- Une molécule appartient à un utilisateur et à des groupes.
- Un utilisateur appartient à des groupes.
- Un groupe a des utilisateurs qui ont le rôle d'administrateur.
- Un groupe est une liste d'utilisateurs ou de groupes.
- La valeur de solubilité d'un solvant a une source bibliographique ou expérimentale.

IV.4.1 Autres spécificités

Ce complément d'informations précise des contraintes dans la réalisation du logiciel.

- Une molécule est définie par un nom, une formule chimique représentée soit par une chaîne de caractères, soit une image et/ou une image en trois dimensions .
- Une méthode de calcul ne peut se faire pour une molécule que si toutes les sous molécules ont les valeurs requises.
- La décomposition n'a pas besoin de prendre en compte l'ordre des sous-molécules, seul compte le nombre de sous-molécules.
- Le calcul ne peut se faire pour une molécule que si toutes les sous molécules ont les valeurs requises.
- La décomposition est de la responsabilité de l'utilisateur
- L'utilisateur est défini, au minimum, par son courriel, un nom et un mot de passe.
- Le groupe est défini par un nom.
- Une solubilité permet de connaître l'affinité entre une substance représentée par la molécule et le solvant.

IV.5 Modèle conceptuel de données

A l'aide des phrases de la partie précédente, nous pouvons construire un modèle conceptuel de données. Cette construction se fait en plusieurs étapes, dont la lecture n'est pas toujours facile. Je propose de présenter un schéma initial, sans aucune optimisation, afin de permettre la visualisation des entités et des relations. Enfin, je proposerai un schéma conceptuel de données.

IV.5.1 Schéma initial conceptuel de données

Pour aider à la lecture, je propose le schéma conceptuel de données, dont les entités et les relations seront présentées aux paragraphes suivants. Un schéma conceptuel de données est fourni après les choix d'optimisation retenus, à la fin de cette sous-partie (voir IV.5.5 Schéma conceptuel de données page 36)

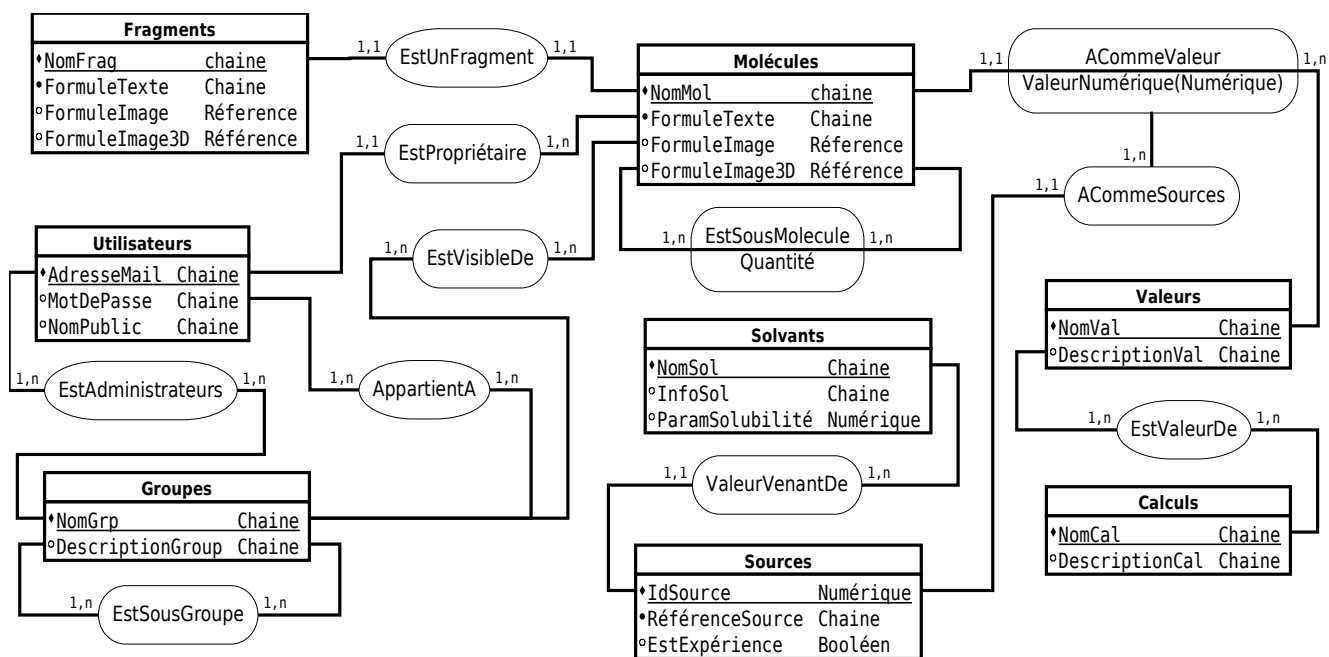


Figure 5 : Schéma conceptuel initial de données

légende : point noir devant les noms des attributs sont pour une valeur non nulle, les entités sont dans des cadres carrés, les relations, dans des cadres arrondis. Les attributs soulignés sont les clefs primaires des entités. Les types de valeurs sont donnés simplement (Chaîne, Référence, Booléen, Numérique). Les liens sont entre entités et/ou relation, la position n'a pas d'importance vis à vis des attributs.

IV.5.2 Entités

Nous pouvons donc observer que fragment, molécule, utilisateur, groupes, sources, calcul et valeur sont des entités. Fragments et molécules possèdent les mêmes propriétés, la première étant un cas particulier de la seconde. Les entités peuvent être décrites comme suit, sans optimisation aucune. Les propriétés soulignées sont les clefs primaires :

Fragments(Nom, FormuleTexte, FormuleImage, FormuleImage3D) où Nom est le nom de référence de la molécule, FormuleTexte correspond à la formule chimique sous forme de texte, FormuleImage est la formule chimique sous forme d'image, FormuleImage3D est la formule en trois dimensions.

Molécules(Nom, FormuleTexte, FormuleImage, FormuleImage3D) où les attributs sont identiques à ceux du Fragments. Cette entité est l'entité principale de la base, autour de laquelle s'articule l'ensemble des relations et des autres entités.

Utilisateurs(Adresse, Mot de passe, Nom, Description) où l'adresse permet d'identifier un seul utilisateur, le mot de passe permet de reconnaître l'utilisateur. Le nom permet de ne pas afficher l'adresse, il ne sera pas forcément unique. l'attribut Description est un champ d'information sur l'utilisateur.

Groupes(Nom, Description) où Nom est le nom du groupe et Description est un champ d'information sur le groupe.

Valeurs(Nom, Description) où le Nom est le nom de la valeur, qui en informatique correspond plus à la notion de variable. La Description permet de connaître l'utilisation ou sa destination. Valeurs contiendra les noms des valeurs et non pas les valeurs numériques, car celles-ci dépendent d'autres entités comme « Molécules ».

Calculs(Nom, Description) où le Nom est le nom du calcul. La Description permet d'avoir un peu plus d'informations sur le calcul.

Sources(ID, Références, Expérience) où ID est un identificateur, Références est un ensemble, sous forme de texte, permettant de retrouver l'information. L'expérience permet de savoir si cela est en rapport avec un résultat obtenu par une manipulation.

Solvants(Nom, Informations, ParamSolubilité) où Nom est le nom du solvant. La propriété Informations aide à donner des informations sur le solvant. ParamSolubilité est la valeur du paramètre de solubilité.

Pour que l'ensemble fonctionne, les entités sont reliées par des relations que j'ai nommées au plus proche de leur utilisation. Elles sont présentées dans la section suivante.

IV.5.3 Relations

Afin d'organiser ces entités, il est nécessaire de poser les relations. Elles sont décrites avec entre parenthèses le nom des tables mises en jeu. Je précise le type de relation et les attributs nécessaires.

EstUnFragment(Fragments, Molécules) relation une pour une. Au vu des entités qu'elle relie, c'est une relation qui qualifie un état : un fragment est une molécule particulière, il fait partie du référentiel.

EstSousMolecule(Molécules, Molécules) relation plusieurs à plusieurs, avec un attribut Quantité pour préciser le nombre de sous-molécules nécessaires. Cette relation représente l'arc entre deux nœuds ou entre un nœud et une feuille de l'arbre de décomposition. Elle ne doit pas être réflexive, même pas transitivité, car cela conduirait à une boucle sans fin d'introspection.

ACommeValeur(Molécules, Valeurs) relation une pour plusieurs, avec un attribut ValeurNumérique représentant la valeur pour la molécule. Il n'y a qu'une valeur numérique pour un couple molécule et valeur.

ACommeSources(ACommeValeur, Source) relation de plusieurs à une, est un cas particulier, car elle représente la source de la donnée numérique. Elle-même dépendante de la relation ACommeValeur, citée juste au dessus.

EstValeurDe(Valeurs, Calculs) relation de plusieurs à plusieurs. Cette relation permet de décrire l'ensemble des variables désignées par l'entité Valeurs qui sont nécessaires aux calculs d'une méthode, représentée par Calculs.

EstPropriétaire(Utilisateurs, Molécules) relation de une à plusieurs. Cette relation permet de désigner l'utilisateur propriétaire de la molécule.

AppartientA(Utilisateurs, Groupes) relation de plusieurs à plusieurs. Donne l'information correspondant à la notion de groupe : quels utilisateurs pour quels groupes.

EstAdministrateurs(Utilisateurs, Groupes) relation de plusieurs à plusieurs. Elle permet de

décrire les utilisateurs administrateurs de groupe vis à vis des groupes.

ValeurVenantDe(Solvants, Sources) relation de plusieurs à un. Elle donne la source unique en relation avec le solvant. La donnée associée au solvant est le paramètre lui-même. Et donc, qui peut nous garantir cette information?

EstSousGroupe(Groupes, Groupes) relation de plusieurs à plusieurs. Un groupe peut être composé de sous groupes, cette relation permet de connaître quels groupes sont associés à quels autres groupes. La relation ne doit pas être réflexive, même par transitivité.

EstVisibleDe(Molécules, Groupes) relation de plusieurs à plusieurs. Représente les groupes qui ont la capacité de pouvoir voir une molécule.

IV.5.4 Optimisations

Cet première étude peut être facilement optimisé. Je propose les optimisations suivantes :

IV.5.4.1) Au niveau des entités

- Le nom d'une molécule ou d'un fragment peut être long et pas toujours simple. Je propose donc d'utiliser un attribut identifiant numérique, plus facile d'utilisation, cela nous donne pour l'entité Molécules :

Molécules(IdMol, Nom, FormuleTexte, FormuleImage, FormuleImage3D)

- De même pour la table Utilisateurs et Groupes, un identifiant unique de forme numérique est souhaitable, nous obtenons ces entités :

Utilisateurs(IDuti, Adresse, mot de passe, nom, DescriptionUtilisateur) où IDuti, est l'identifiant numérique.

Groupes(IDgrp, Nom, DescriptionGroup) où IDgrp est l'identifiant numérique.

IV.5.4.2) Au niveau des relations

- La relation simple **EstUnFragment** peut devenir un attribut de l'entité Molécules, permettant de fusionner les deux entités Molécules et Fragments. De plus, la relation EstPropriétaire peut être facilement réduite en un attribut qui s'ajoute à Molécules.

Molécules(IdMol, Nom, FormuleTexte, FormuleImage, FormuleImage3D, EstUnFragment, EstPropriétaire) où *EstUnFragment* correspond à une valeur binaire (Oui/Non) et *EstPropriétaire* est un attribut ayant comme valeur l'identifiant de l'entité Utilisateurs (IDuti).

- La relation **ACommeSources** est simplifiable en la mettant comme attribut de la relation **AcommeValeurs**, elle contiendra l'identifiant de l'entité Source.
- **ACommeValeur**(Molécules, Valeurs) relation une pour plusieurs, avec deux attributs *ValeurNumérique* représentant la valeur pour la molécule et *AcommeSource*, ayant l'identifiant de l'entité Sources correspondante.

IV.5.5 Schéma conceptuel de données

Le schéma conceptuel de données ci-après est un schéma incluant les différentes optimisations décrites ci-avant.

Ainsi, certaines relations n'apparaissent plus, transformées en attributs. Ce schéma est plus proche d'un schéma physique des données, mais d'autres optimisations seront présentées plus loin, car difficilement représentables dans un schéma conceptuel.

Les cardinalités des liaisons sont données différemment. Durant l'optimisation, certaines relations sont transformées en de simples attributs intégrés dans une entité, réduisant le schéma. Les liaisons sont maintenant définies entre attributs, sinon le schéma perdrait certaines relations entre deux entités.

Pour des raisons de clarté, j'ai modifié la position de certaines tables vis à vis la « Figure 5 : Schéma conceptuel initial de données. » page 32.

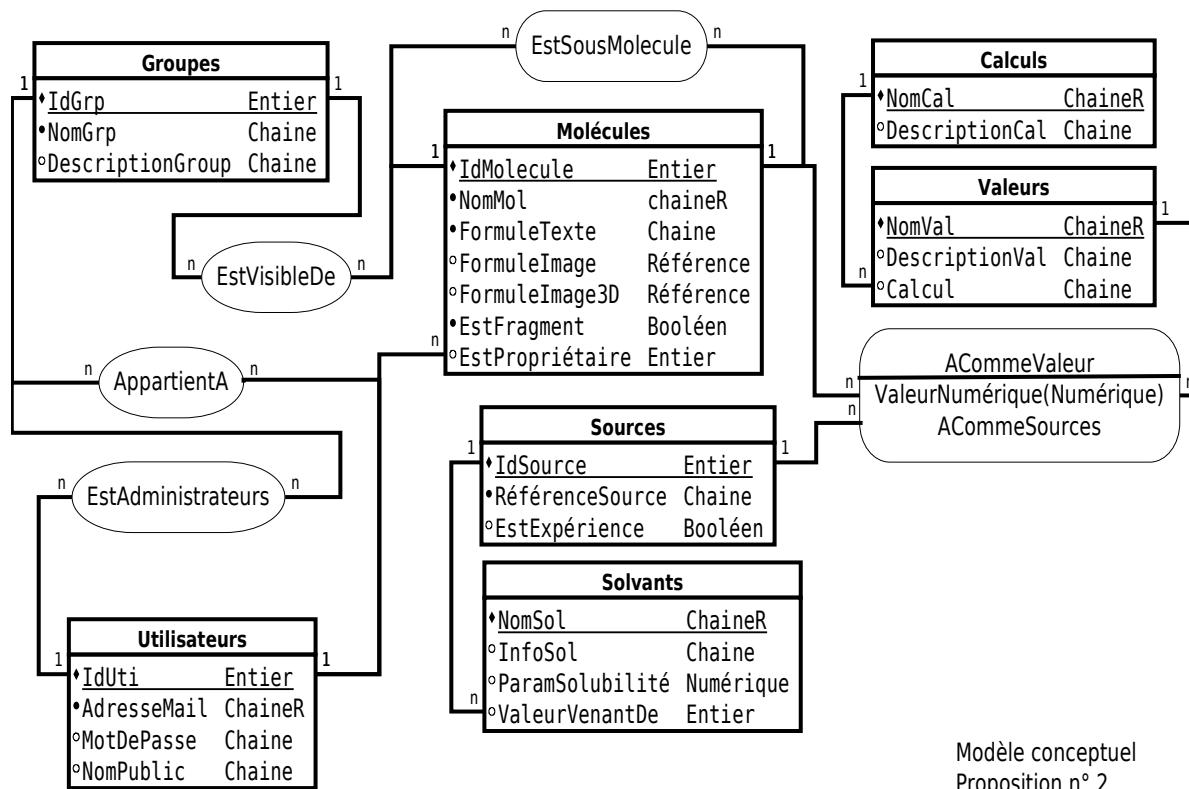


Figure 6 : Schéma conceptuel de données après les premières optimisations

légende : point noir devant les noms des attributs sont pour une valeur non nulle, les entités sont dans des cadres carrés, les relations, dans des cadres arrondis. Le type de valeurs qui sont données sont un peu plus complexes. Deux types de chaînes, l'une représente une chaîne dont nous pouvons faire une recherche notée **ChaineR**, l'autre est un champ de caractères et il est noté **Chaine**. Le type de valeur **Entier** correspond à des valeurs numériques entières qui peut être de grand nombre et Numérique correspond à des valeurs décimales. **Référence** est une méthode pour permettre de stocker un lien, interne ou externe, vers des informations complexes comme une image. Ici les liens sont positionnés sur les attributs, optimisation oblige.

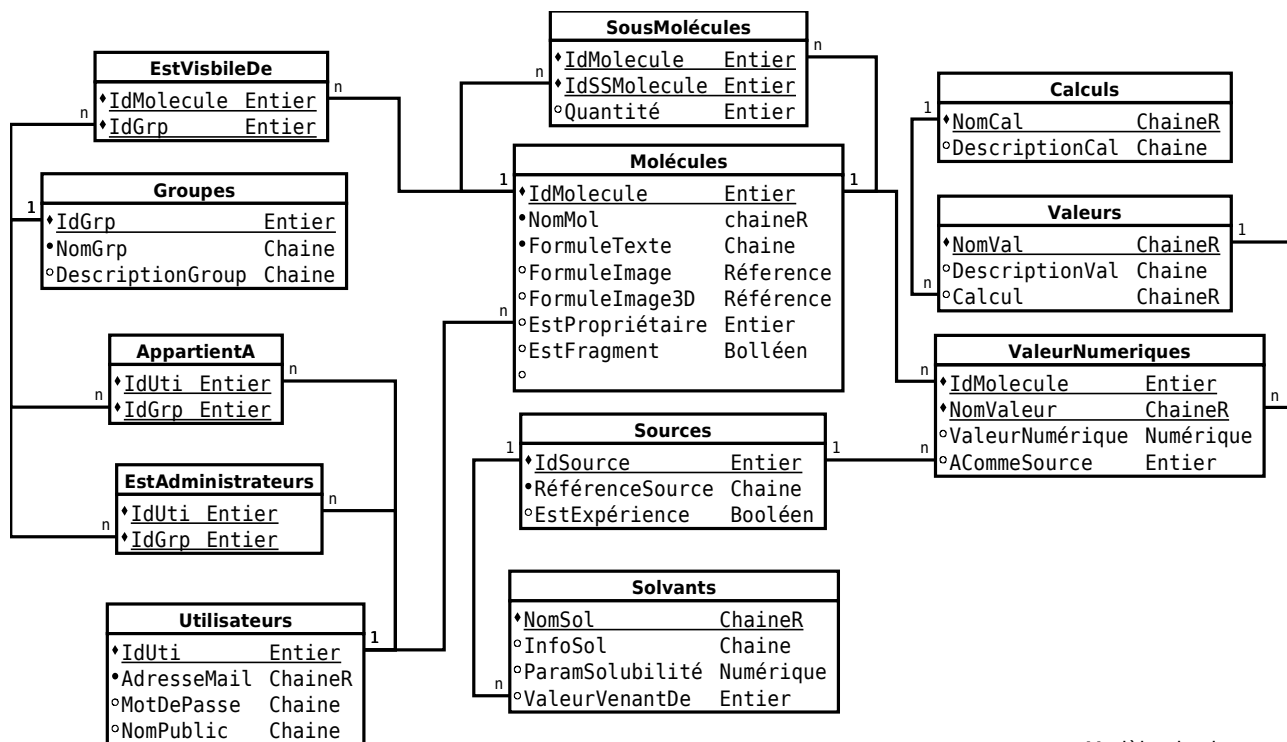
L'étape suivante est la projection de toutes les entités et toutes les relations sous forme de tables. Ce qui donnera un modèle physique des données.

IV.5.6 Modèle physique des données

N'ayant pas encore choisi la base de données, ni la plateforme de développement, le modèle physique des données proposé reste assez théorique. Dans la partie VI. Réalisation à la page 67, je proposerai le modèle utilisé pour le développement.

La transformation des entités vers les tables est triviale, chaque attribut correspond un champs et les champs servant de clef primaire ont été soulignés.

Les relations ont été transformées en table en prenant les champs correspondants aux clefs primaires de chaque table. On peut remarquer que la table ValeurNumériques ne reprend que les clefs primaires de Molécules et Valeurs. Nous pouvons aussi remarquer que la table SousMolécules, utilisant deux fois la même table, j'ai changé le nom de la clef pour rester cohérent avec un nom de champ unique par table.



Modèle physique

Figure 7 : Modèle physique de données

légende : point noir devant les noms des attributs sont pour une valeur non nulle. Les tables sont les cadres avec leurs noms dans la partie supérieure, les champs dans la partie inférieure, les champs soulignés correspondent aux clefs primaires. Deux types de chaînes, l'une représente une chaîne dont nous pouvons faire une recherche notée **ChaîneR**, l'autre est un champ de caractères et il est noté Chaîne. Le type de valeur **Entier** correspond à des valeurs numériques entières, qui peut être de grand nombre et **Numérique** correspond à des valeurs décimales. **Référence** est une méthode pour permettre de stocker un lien, interne ou externe, vers des informations complexes comme une image. Ici les liens sont positionnés sur les champs.

Après l'analyse des besoins et l'analyse des données, je vous présente l'analyse que j'ai réalisée sur les traitements. Mais tout d'abord, pour une meilleure compréhension, je cite dans le sous-chapitre suivant le paramètre de solubilité tel que J. Revilla le décrit dans sa thèse [RJ94] et j'expose des méthodes de calculs pour l'obtenir. Puis je propose de reprendre le cahier des charges pour en déduire les traitements à réaliser.

IV.6 Paramètre de solubilité

Le concept de paramètre de solubilité développé par Hildebrand[HJ49] a été introduit dans la théorie des solutions. En effet, tout liquide pur est caractérisé par une énergie de cohésion E_{coh} , qui correspond principalement à son énergie molaire de vaporisation lorsque le gaz résultant est à pression nulle.

$$E_{coh} = \Delta H_v - P \Delta V = \Delta H_v - RT$$

où : ΔH_v = enthalpie de vaporisation

ΔV = variation de volume

P = pression

R = la constante des gaz c'est à dire $1.987 \text{ Cal.mol}^{-1}.\text{K}^{-1}$

T = la température en Kelvin.

La densité d'énergie de cohésion $e = E_{coh}/v$, représente de l'énergie totale nécessaire pour amener les molécules contenues dans le volume molaire v , à une distance infinie les unes des autres. Le paramètre de solubilité δ est défini comme la racine carrée de la densité d'énergie de cohésion e .

Par ailleurs, l'énergie libre de mélange de deux composés (ΔG_m), est donnée par la relation :

$$\Delta G_m = \Delta H_m - T \Delta S_m$$

où : ΔH_m = enthalpie de mélange

ΔS_m = entropie de mélange

Il y a solubilisation quand $\Delta G_m < 0$, c'est à dire quand l'enthalpie de mélange ΔH_m est inférieure au terme entropique $T \Delta S_m$ (positif). La théorie des solutions permet d'exprimer ΔH_m en fonction des paramètres de solubilité des liquides (1 et 2) purs δ_1 et δ_2 . Les hypothèses sont les suivantes :

- lors du mélange, le changement de volume est nul
- l'énergie d'interaction entre une molécule du liquide 1 et une molécule du liquide 2 est

la moyenne géométrique des densités de cohésion des deux liquides correspondants :

$$e_{1,2} = \delta_1 \cdot \delta_2 \text{ avec } \delta_1 = (e_1)^{1/2} \text{ et } \delta_2 = (e_2)^{1/2}$$

En supposant que la contribution de chacun des composés à la densité totale d'énergie de cohésion du mélange est proportionnelle à sa fraction volumique Φ , l'expression de l'enthalpie est donnée par la relation de Hildebrand[VK76] suivante :

$$\Delta h_m = \Phi_1 \Phi_2 (\delta_1 - \delta_2)^2$$

avec : Δh_m = enthalpie de mélange par unité de volume

Φ_1 et Φ_2 = fractions en volume des composants 1 et 2

δ_1 et δ_2 = paramètre de solubilité des composants 1 et 2

ΔH_m est donc toujours positif ; il est d'autant plus faible que les paramètres de solubilité des deux composés en présence sont proches. Les conditions optimales de solubilisation sont obtenues quand $\delta_1 = \delta_2$ ce qui entraîne $\Delta H_m = 0$.

En fait, l'équation de **Hildebrand** ne tient compte que des forces de dispersion entre les molécules, elle n'est donc valable que pour des composés non polaires. Cependant, pour de nombreux composés, l'énergie de cohésion dépend aussi des interactions entre les groupes polaires, et les liaisons hydrogènes. Hansen a montré que l'énergie de cohésion peut être décomposée en trois parties correspondant aux trois types de force attractives :

$$E_{\text{coh}} = E_d + E_p + E_h$$

où E_d = contribution des forces de dispersion de London qui représentent les interactions entre dipôles induits

E_p = contribution des forces de Keesom, qui représentent les interactions entre dipôles permanents

E_h = contribution des forces dues aux liaisons hydrogènes

D'où, l'équation correspondante pour le paramètre de solubilité qui fait apparaître des paramètres de solubilité partiels :

$$\delta^2 = \delta_d^2 + \delta_p^2 + \delta_h^2$$

L'enthalpie de mélange par unité de volume Δh_m s'écrit alors :

$$\Delta h_m = \Phi_1 \Phi_2 [(\delta_{d1} - \delta_{d2})^2 + (\delta_{p1} - \delta_{p2})^2 + (\delta_{h1} - \delta_{h2})^2]$$

Les paramètres de solubilité partiels d'un composé peuvent être déduits de manière théorique de sa structure chimique, à partir des contributions des différents groupes qui la constituent :

$$\delta_d = \frac{\sum F_{di}}{V_a} \quad \delta_p = \frac{\sqrt{\sum F_{pi}^2}}{V_a} \quad \delta_h = \frac{\sqrt{\sum E_{hi}}}{V_a}$$

F_{di} et F_{pi} sont des constantes d'attraction molaire, caractéristiques des différents groupes entrant dans la constitution chimique d'un composé. E_{hi} correspond à l'énergie des liaisons hydrogènes par groupe structural.

Dans le tableau 1 sont reportées les valeurs F_{di} , F_{pi} , E_{hi} ainsi que les volumes molaires partiels v_i des différents groupements chimiques entrant dans la structure du composé saccharidique LIMA (exemple) [VK76]

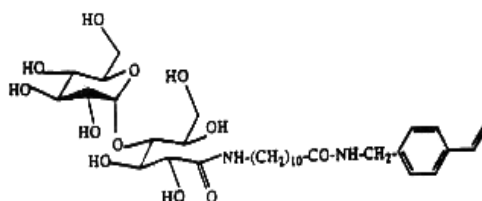


Figure 8: la représentation chimique de la molécule LIMA

groupe chimique	v_i ($\text{cm}^3 \cdot \text{mole}^{-1}$)	F_{di} ($\text{J}^{1/2} \cdot \text{cm}^{3/2} \cdot \text{mole}^{-1}$)	F_{pi} ($\text{J}^{1/2} \cdot \text{cm}^{3/2} \cdot \text{mole}^{-1}$)	E_{hi} ($\text{J} \cdot \text{mole}^{-1}$)
=CH2	20,51 (C=C)	400	0	0
=CH-		200	0	0
	62,4	1270	110	0
CH2	15,8	270	0	0
NH	8,57	160	210	3100
C=O	15,18	290	770	2000
CH	12,33	80	0	0
-OH	6,74	210	500	20000
-O-	7,28	100	100	3000

Tableau 1: Volumes molaire partiels et contributions aux paramètres de solubilité partiels des différents groupes chimiques

Le calcul du paramètre de solubilité par cette méthode, donne pour la molécule LIMA :

$$\delta = 12,5 \text{ (cal.cm}^{-3}\text{)}^{1/2}$$

Cette valeur peut être comparée aux paramètres de solubilité de différents solvants donnés dans ce tableau :

Solvants	$\delta \text{ (cal.cm}^{-3}\text{)}^{1/2}$
Chloroforme	9,3
Acétone	9,9
DMSO	12,0
DMF	12,1
Méthanol	14,5
Eau	23,4

Tableau 2: Paramètre de solubilité de quelques solvants

Il apparaît clairement que le paramètre de solubilité du monomère saccharidique est en accord avec les résultats expérimentaux des essais de solubilité, puisqu'il est très soluble dans le diméthylformamide (DMF) et diméthylsulfoxyde (DMSO). Le monomère LIMA est légèrement soluble dans le méthanol (environ 10 g/l), mais insoluble dans le chloroforme et dans l'acétone.

Après avoir apporté une présentation du paramètre de solubilité, je vous expose les différentes méthodes qui m'ont été fournies et qui devront être intégrées à plus ou moins long terme dans le programme.

IV.7 Différentes méthodes de calculs

L'utilisateur va décomposer sa structure moléculaire en fonction de sous molécules disponibles dans la base de données des valeurs, allant jusqu'aux fragments, qui peuvent être de simples atomes, décrits dans un livre [VK09]. Hélas, les valeurs dont le volume, la liaison

hydrogène, la force de London et celle de Keesom, ne sont pas souvent regroupées ensemble ou dépendent du reste de la composition. Ces valeurs dépendent aussi de la méthode choisie, c'est à dire que le calcul peut être vu sous différents angles, sans qu'il soit possible de passer d'une méthode à une autre.

δ_D : la force de dispersion, correspondant à la force de London, est la somme des énergies qui poussent les atomes à la dispersion.

δ_P : la force de la polarisation, correspond à la force de Keesom, est la somme des énergies qui tendent à une forme d'organisation.

δ_H : la force de liaison hydrogène ou liaison covalente, est la somme des énergies qui tendent au maintien de la structure.

Le volume relatif a une part importante dans les calculs, suivant le type de fragment, une molécule peut avoir un volume plus faible ou plus réduit que le volume de chacun de ses fragments.

IV.7.1 Calcul par la méthode de Van Krevelen

δ_D = la somme des force de dispersion sur le volume relatif.

δ_P = la racine carrée de la somme des forces de Keesom, au carré, le tout sur le volume relatif.

δ_H = la racine carrée de la somme des forces de liaison divisée par le volume relatif.

$$\delta = \sqrt{\delta_D^2 + \delta_P^2 + \delta_H^2}$$

$$\delta_D = \frac{\sum F_{di}}{V_a} \quad \delta_P = \frac{\sqrt{\sum F_{pi}^2}}{V_a} \quad \delta_H = \frac{\sqrt{\sum E_{hi}}}{V_a} \quad V_a = \sum V_{ai}$$

les valeurs de F_{di} , F_{pi} , E_{hi} sont référencées dans le tableau 7.10 du livre [VKR1] donné en annexe dont une petite partie est retranscrite ci dessous :




Groupe Structurel	F_{di} (MJ/m ³) ^{1/2} . mol ⁻¹	F_{pi} (MJ/m ³) ^{1/2} . mol ⁻¹	E_{hi} J/mol
-CH ₃	420	0	0
-CH ₂ -	270	0	0
>CH-	80	0	0
>C<	-70	0	0
=C<	70	0	0
	1620	0	0
	1430	110	0
 (o, m, p)	1270	110	0
-F	(220)	-	-
-Cl	450	550	0
-Br	550	-	-
-CN	430	1100	2500
-OH	210	500	20000
=PO	740	1890	13000
Anneaux	190	-	-
• Symétrie sur un plan	-	0,5 x	-
• Deux plans de symétrie	-	0,25 x	-
• Plusieurs plans de symétrie	-	0 x	0x

Tableau 3: Extrait du tableau des contributions des groupes de composants aux paramètres de solubilité (Van Krevelen)

Nous pouvons observer dans le tableau-ci dessus que certains fragments n'ont pas de valeur ou leur valeur est à nulle, indiqué par le signe « - », interdisant le calcul. D'autres valeurs sont à zéro, différentes de la valeur nulle et permettant le calcul. Nous pouvons aussi remarquer que des valeurs dépendent des fragments de molécule intervenant dans la constitution de la molécule comme les anneaux. Même les liaisons de covalence sont importantes, par exemple >C< et =C< qui possède de F_{di} dont les valeurs sont négatives l'une

par rapport à l'autre.

Par ailleurs, les valeurs de V_{ai} sont référencées dans le tableau 4.10 du même livre [VK07] donné en annexe et dont un extrait pour l'exemple est fourni à la ci-dessous :


Groups (Bivalent)	$V_{a,i(298)}$ cm ³ /mol	$V_{c,i(298)}$ cm ³ /mol	Groups	$V_{a,i(298)}$ cm ³ /mol	$V_{c,i(298)}$ cm ³ /mol
-CH ₂ -	16,37	14,68	-CO-al	13,5	12,25
-CH(CH ₃)-	32,72	29,35	-CO-ar	18,5	16,8
-CH(t-C ₄ H ₉)-	65,44	58,69		86	(77)
-CH(C ₆ H ₁₁)	101,76	91,27	-CH(C ₆ H ₅)-	84,18	75,48

Tableau 4: Extrait du tableau des valeurs recommandées pour le volume molaire

A partir de cet extrait, nous remarquons que les fragments sont plus complexes que dans le Tableau 3 précédent. Ainsi pour son utilisation, le chimiste doit décomposer la molécule et la rechercher dans le Tableau 3. Par exemple :

-CH(CH₃)- en ligne 2 du Tableau 4 dont la $V_{a,i}$ est de 32,72 correspond à la décomposition de >CH- + -CH₃, ligne 3 et 2 du Tableau 3.

Nous pouvons aussi observer, que comme pour le Tableau 3, certains fragments dépendent des autres fragments, comme par exemple, -CO-ar, qui demande d'avoir un fragment aromatique, les autres cas devant utiliser la ligne juste avant correspond au fragment -CO-al.

IV.7.2 Calcul par la méthode de Hoy

Le problème de la méthode vient du fait que certaines valeurs sont approximatives, pour faciliter les calculs et que, à contrario de la méthode de Van Krevelen, les valeurs sources (F_p , F_t) sont connues sur un ensemble plus restreint de molécules.

$$\delta = \sqrt{\delta_t^2 - \delta_p^2 - \delta_h^2}$$

$$\delta_t = \frac{\sum F_t + \frac{B}{n}}{\sum \gamma} \quad \delta_p = \delta_t \left(\frac{1}{\alpha^2} \frac{\sum F_p}{\sum F_t + \frac{B}{n}} \right)^{1/2} \quad \delta_h = \delta_t \left[\frac{(\alpha^{(p)} - 1)}{\alpha^{(p)}} \right]^{1/2}$$

B = base value, une constante de valeur 277

$$n = \frac{0,5}{\sum \Delta_t^{(p)}} \quad \alpha^{(p)} = \frac{777 \sum \Delta_t^{(p)}}{\sum \gamma}$$

F_p, F_t, Δ t, γ sont données dans le tableau 7.12 du livre de référence [VK09]

IV.7.3 Calcul par la méthode de Small

Cette méthode à l'image de celle de Hoy, utilise, en partie, d'autres paramètres. Cette méthode est utilisée principalement pour les macromolécules, en particulier pour les polymère. V_a est la même valeur que celle utilisée pour la méthode de Van Krevelen.

$$\delta = \frac{P_{(P_a)} * \sum F}{M_{w(\text{unité structurelle})}} \quad P_{(P_a)} = \frac{M_{w(\text{unité répétitive})}}{V_a}$$

$$V_a = \sum V_{ai} \quad M_w = \sum M_{atome}$$

F est référencé dans le tableau 2.2 du livre de référence [VK09]

M_{atome} est référencé dans le tableau 4.1 du livre de référence [VK09]

V_{ai} est référencé dans le tableau 4.10 du livre de référence [VK09].

Premières difficultés :

Afin de mener à bien l'ensemble des calculs, il faut que le choix de décomposition arrive à des fragments dont le référentiel possède les différentes valeurs, sinon le calcul ne peut être réalisé.

En supposant que les valeurs du référentiel initial soient de type expérimental, plus la décomposition est grande, plus le calcul devient approximatif et théorique. De plus, la force de Debye étant souvent ignorée, trop faible par rapport aux autres, et les autres valeurs approximatives, plus l'arbre de décomposition est grand, plus le calcul tend à être faussé. Il est donc intéressant de connaître le pourcentage d'approximation ou la profondeur de l'arbre avant une feuille (un fragment du référentiel).

Ainsi, une molécule composée uniquement de fragment du référentiel, a une profondeur de 1, correspondant à la meilleure approximation possible.

IV.8 Respect des contraintes du cahier des charges

Durant la première phase du projet un cahier des charges a été élaboré, dont le texte ci-dessous donne les grandes lignes.

IV.8.1 Cahier des charges

Afin de compléter l'étude, je vous présente les parties du cahier des charges. Je reprends l'ensemble des points, ceci afin de vérifier que je réponde à l'ensemble des fonctionnalités et des contraintes exprimées. Je commence par rappeler l'objectif du logiciel, ceci dans l'introduction du document, puis je donnerai les fonctionnalités qui ont été demandées, suivies des contraintes associées.

IV.8.1.1) Introduction

L'introduction du document permet de poser et de comprendre la première problématique, du point de vue du chercheur :

« Pour un produit rare et non commercial, faire des tests expérimentaux dans différentes solutions peut devenir très rapidement cher et long. Un étudiant produisant au bout d'un an et demi, après 17 étapes d'élaboration, quelques milligrammes voir quelques grammes d'un produit actif, ne peut envisager de les perdre dans une solution, sans savoir si le produit et la solution ont le comportement souhaités. »

La notion de non commercial est celle de la réalisation par le chercheur d'un produit ne pouvant être obtenu par l'achat, ce qui est par définition un produit rare. Ce texte n'est pas assez précis, mais il a permis de poser certaines fonctionnalités, dont voici la liste.

IV.8.1.2) Fonctionnalités

Je donne la liste des fonctionnalités que je reprends les unes après les autres pour les détailler à la suite :

1. Le logiciel doit calculer le paramètre de solubilité d'une molécule.
2. La saisie de fragments moléculaires contient la possibilité d'images symboliques
3. Un indicateur pour savoir si le résultat est expérimental ou calculé.
4. L'exposition des calculs, avec une partie de la résolution chiffrée, afin de permettre la mise au point et la validation.
5. La saisie des solvants (ou milieux) est possible.
6. La présentation des résultats et des choix des solvants doit être affichée.
7. La minimisation des saisies (retour aux modifications) doit être recherchée.
8. La sauvegarde de molécules avec leur paramètre de solubilité doit être possible.
9. La diffusion de l'application doit être facilitée : elle jouera le rôle d'une vitrine du Laboratoire.
10. La disponibilité de l'application est demandée.
11. La maintenance du logiciel et du référentiel devra être prévue.

Pour plus de clarté, Je reprends cette liste que je détaille par ligne ou groupe de lignes ci-dessous

1) Le logiciel doit calculer le paramètre de solubilité d'une molécule

La première fonctionnalité étant le cœur du logiciel, et exposé précédemment, je ne donnerai pas plus d'information ici.

2) la saisie de fragments moléculaires contient la possibilité d'images symboliques

Le second point est prévu dans la structure, une image pourra être associée à une molécule et le logiciel devra prendre en compte les contraintes associées que sont l'utilisation de gros objets, la possibilité de télécharger une image pour l'intégrer à la base de données.

3) Un indicateur pour savoir si le résultat est expérimental ou calculé

Le troisième point est la formulation d'une demande pour avoir une échelle qui mesure la distance entre l'expérimentation ou les valeurs connues de façon certaine vis à vis de la valeur obtenue. L'objectif est de connaître le degré d'éloignement entre les fragments et la molécule souhaitée. Chaque niveau de profondeur supplémentaire dans l'arbre de décomposition, rend le résultat plus théorique.

4) L'exposition des calculs, avec une partie de la résolution chiffrée, afin de permettre la mise au point et la validation.

5) La présentation des résultats et des choix des solvants doit être affichée.

Les points 4 et 5 se rejoignent. Le chimiste doit pouvoir vérifier ses propres résultats et calculs. lui permettre de connaître la décomposition exacte qui a été choisie pour sa molécule et d'avoir les calculs intermédiaires pour voir quelle pourrait être la source d'une différence. En effet, le choix d'une molécule comme sous-molécule implique une potentielle influence, espérée minime dans le résultat.

6) La saisie des solvants (ou milieux) est possible.

La saisie des solvants permettra d'ajouter et de gérer d'autres solvants, qui ne sont pas forcément dans le domaine dans lequel le LAGEP travaille; ouvrant à plus de possibilités. Il faudra marquer les résultats afin de permettre de connaître les différences types sources possibles. Il est préférable de distinguer entre les sources fournies par la bibliographie et celle par l'expérimentation. C'est aussi valable pour savoir si le solvant provient du référentiel.

7) La minimisation des saisies (retour aux modifications) doit être recherchée

Le point sept peut être interprété sur différents axes. Le premier axe est la capacité à reprendre des valeurs déjà saisies pour les utiliser dans une construction plus complexe : typiquement celle des molécules utilisant des molécules déjà existantes comme sous-molécules. Le second axe est celui de la possibilité d'utiliser une même molécule, mais de l'enregistrer sous une variation, soit pour modifier une de ses décompositions soit une de ses valeurs, dans le cas d'un fragment. Le troisième axe porte sur la visibilité des molécules vis à vis d'autres utilisateurs que celui qui les a créées. Ici, l'utilisateur aura la possibilité de fournir à un groupe la vision d'une molécule. Nous sommes dans la demande de gestion des groupes et des droits liés aux groupes et aux utilisateurs.

9) La diffusion de l'application doit être facilitée : elle jouera le rôle d'une vitrine du Laboratoire.

Nous souhaitons mettre à disposition de l'ensemble de la communauté un logiciel gratuit. Alimentée par les chercheurs et utilisateurs, cette base de données deviendra une référence, gage de qualité.

10) La disponibilité de l'application est demandée.

Ce point exprime l'envie d'avoir une application non seulement accessible ou utilisable par l'ensemble des chercheurs intéressés, sans forcément avoir à s'identifier. Ceci permettant une facilité d'accessibilité de n'importe quel endroit sur la planète. Au besoin, l'application peut être mise sur un poste non connecté ou être fournie pour un autre laboratoire avec seulement le référentiel. L'application travaillerait dans ce cas sur un espace plus réduit de molécules ; seules celles qui ont été saisies sur le poste ou dans le laboratoire enrichiront le

catalogue de l'application.

11) La maintenance du logiciel et du référentiel devra être prévue

Ce point implique que l'application devra pouvoir évoluer sans que l'utilisation en soit perturbée. Ainsi, la mise au point, le débogage ou la création de nouvelles fonctionnalités aura un impact le plus réduit possible par rapport à l'existant.

IV.9 Choix des plateformes

Dans cette partie du document, je vais présenter les différentes solutions envisageables, leurs avantages et leurs inconvénients, de mon point de vue. Dans un premier temps, j'opposerai deux systèmes d'exploitation utilisés dans le laboratoire. Dans un second temps, les différentes possibilités entre les langages de programmation avec leur environnement de développement seront exposées. Dans un troisième temps, je comparerai les différentes bases de données disponibles. Puis, je proposerai la solution retenue en mettant en avant ses points forts.

IV.9.1 Windows Serveur ou Linux Serveur

La première question pourrait être : quel système d'exploitation ? Beaucoup de systèmes sont envisageables. Certains peuvent limiter le choix des langages de programmation ou des environnements de développement. Comme je l'ai précisé dans le contexte, au LAGEP, nous avons des serveurs Linux. J'ai aussi la possibilité d'installer un serveur Windows (version 2003 Server). Les autres systèmes d'exploitation peuvent avoir des facilités ou des spécificités intéressantes, comme une interface conviviale pour les serveurs Apple, ou un taux d'attaque plus réduit comme les serveurs openBSD. Mais pour permettre une diffusion la plus large possible, un choix s'impose : soit l'environnement système Windows, soit l'environnement Linux.

Les systèmes Windows sont les systèmes d'exploitation les plus utilisés tant aux

niveaux des clients, autour de 95% [PW5], que des serveurs, avec une part de marché financière suivant IDC, de l'ordre de 35% [PW6]. Ils ont des avantages dont leur utilisation dans les entreprises, ce qui les favorise souvent par rapport aux autres systèmes d'exploitations. La force de vente et l'existence d'une très grosse entreprise rassurent les professionnels. La richesse des applications disponibles, aussi bien pour les environnements de développement, que des instruments de mesures ainsi que pour tous les autres domaines, est un avantage qui souvent l'emporte sur les autres systèmes. Ses points faibles sont liés à la partie financière, le coût des licences, la limitation d'utilisation des licences, l'assistance très vite payante. Les applications développées sur ces systèmes possèdent une complexité, souvent masquée par l'aide au développement et cachant des lourdeurs de programmation, qui donnent l'impression de lenteur ou de perte de puissance. Les paramétrages des applications sont souvent cachées, difficilement accessibles, renforçant le sentiment d'opacité du système. Enfin, étant les systèmes d'exploitation les plus courants, ils sont le terrain d'étude des programmeurs mal intentionnés et la cible privilégiée de nombreux virus.

Les systèmes Linux, et par extension les systèmes Unix, ont la réputation d'être stables, peu consommateurs de ressources. Certaines applications serveurs ont été créées pour ces plate-formes. Les licences sont généralement peu limitatives, la documentation gratuite, mais le manque d'assistance et d'entreprises de références peuvent freiner son utilisation. De plus, il existe des multitudes de distributions avec beaucoup de différences dans les outils de gestion et d'organisation courante. En contre partie, les systèmes Linux ont une philosophie de l'accessibilité, à travers des applications OpenSource[PW7]. C'est à dire, que le code source est disponible et vérifiable voire modifiable. Des fichiers de paramétrage en format texte sont assez simples à lire et à écrire. Par ailleurs, les applications OpenSource sont souvent disponibles et portées sur les autres systèmes d'exploitations.

Linux me paraît donc un bon choix dans l'optique d'une distribution du produit sans contrainte de licence. De plus, les serveurs à ma disposition pour réaliser ce développement sont de ce type.

IV.9.2 Langages de développement

Après le choix du système, j'expose différents critères pour déterminer le langage de programmation. Pour mener à bien cette réalisation, j'ai des notions assez poussées dans le C et C++, le Basic, plus exactement le Visual Basic 6.0 de Microsoft, le JAVA et le PHP avec un peu de JavaScript.

L'application devra avoir une interface graphique, le C et le C++ ne facilitent pas ce genre de développement, n'ayant pas de routine ou d'API graphique définie par défaut. Leur avantage est la rapidité d'exécution, ce qui n'est pas un critère important. Le Basic dans sa version Visual Basic 6, élimine cet inconvénient. Il possède beaucoup de facilité de programmation, avec un environnement de développement complet et pratique à l'usage, comme Java, qui possède des API standards pour les interfaces graphiques. Java étant multi-plateforme, son principal défaut est que je ne possède pas d'environnement de développement me permettant de dessiner rapidement une maquette. Pour ce qui est de Visual Basic, la solution tout Microsoft peut ne pas être en accord avec certains chercheurs, elle nécessite d'utiliser des moteurs de base de données de type « Jet » ou SQL Server, très propriétaire.

De plus, il faut prévoir une base de données qui n'est pas systématiquement sur le poste, ce qui nécessite une configuration, qui est rarement automatisable. Dans ce cas, il faudra, pour tous les langages « compilables » comme Java, VB, et le C++, un module de connexion et l'utilisation soit de technologies propriétaires soit des technologies généralistes, comme ODBC, qui peuvent être un goulot d'étranglement dans l'accès aux données. L'installation est moins aisée donc.

Enfin, le PHP et JavaScript, nécessitent une infrastructure de serveurs, un serveur web avec le langage PHP installé et configuré. Le JavaScript permettant de dynamiser la page, sans forcer le rechargement complet de celle-ci, pratique pour permettre d'avoir une vue plus large que ce que permet une page de type web. L'avantage est la diffusion et l'installation, qui deviennent très simple. Le client n'ayant pas besoin de préparation ou de droits spécifiques,

car il utilise son navigateur. En contre partie, il faut soit éviter les spécificités des navigateurs soit prendre en compte le maximum de navigateurs pour implémenter certaines fonctionnalités. Le langage PHP intègre les éléments nécessaires pour la programmation orienté objet depuis sa version 4, mais celle-ci manquait de maturité. La version 5 semble être plus aboutie, mais je n'ai pas eu le temps de faire des tests. Pour moi, le PHP reste un langage procédural, car chaque page est un programme qui est réinterprété à chaque appel.

De plus, j'ai la charge de plusieurs sites web dynamique, cela nécessite de maintenir et faire progresser mon niveau en PHP. C'est donc naturellement que je choisis le langage PHP, en accord avec l'équipe de travail de monsieur Elaissari.

IV.9.3 Serveurs de base de données

Le langage étant défini, je peux maintenant choisir ce qui pourra gérer les données. Il faut d'abord définir le dimensionnement et les limites que nous souhaitons apporter à nos données.

L'application aura moins d'une centaine de fragments. Donc au mieux, nous avons estimé un millier de fragments disponibles, c'est à dire que les chercheurs du laboratoire ou de confiance donneront assez d'informations, de valeurs et de preuves pour valider un millier de molécules que nous intégrerons au référentiel.

D'un autre coté, les calculs et leurs variables sont de faibles unités. Nous comptons sur 3 méthodes, chacune ayant jusqu'à 4 variables. Ce qui nous fait moins d'une vingtaine de valeurs.

Il est souhaitable que la table des utilisateurs grossisse régulièrement. De plus, une personne peut avoir plusieurs comptes. Actuellement, il n'y a pas de procédure de purge ou d'élimination des comptes. Donc, la table ne peut que grossir. Si nous atteignons 1000 comptes, nous aurons réussi notre objectif de diffusion. Par sécurité, nous pouvons fournir le nombre de 100 000 comptes, mais déjà à ce niveau, nous rencontrerons d'autres problèmes liés aux chaînons de notre infrastructure entre nous et l'université.

La table qui générera beaucoup de données est celle des molécules. Sa taille dépendra du nombre d'utilisateurs, des calculs et de leurs valeurs (les variables de calculs), et surtout du nombre moyen de molécules créées et produites par les utilisateurs. Ce nombre peut grandir rapidement. Si on suppose qu'un chercheur chimiste travaille avec une centaine de molécules par an, sur 40 ans de carrière, nous obtenons 4 000 molécules par compte. Avec un nombre de comptes estimé à 100 000, nous obtenons 400 000 000 de lignes ou quatre cents millions de lignes d'enregistrement. Ceci nous permet d'exclure les petits moteurs de base de données, comme SQLite[PW8] ou bien une base Microsoft Access et son moteur JET. Ils peuvent quand même servir pour aider à réaliser une maquette.

Il existe plusieurs systèmes de gestion de base de données dont j'ai, soit des notions, soit des utilisations poussées. Durant mes activités professionnelles dans le secteur privé, j'ai travaillé avec SQL Server de Microsoft, et Oracle. De plus, du côté Open Source, le choix est entre Postgresql et Mysql.

Pour SQL Server [PW9], celui-ci nécessite un serveur Windows pour tourner. Une version gratuite existe désignée sous le nom de SQL Server version express. SQL Server se veut être le moteur de base de données le plus polyvalent possible en terme d'accès et de dimension des données. La version express possède des limitations comme un seul processeur et un maximum de 1 Go de mémoire RAM.

Oracle est moins contraignant sur le type de serveur, il accepte de s'installer sur des environnements Linux et Windows. Oracle nécessite une grosse infrastructure machine vis à vis des données que le projet pourrait avoir à gérer, même s'il propose une version plus légère : « Oracle Database Express Edition » [PW10]. Cette version gère actuellement des fichiers de données limités à 4 Go de données et un serveur doté d'un seul processeur et d'un maximum de 1 Go de mémoire RAM.

Du côté de l'Open Source, Postgresql[PW11] est un système de gestion de base de données dont je ne dispose pas de beaucoup d'outils. Je le connais que superficiellement. Comme les autres SGBD, il a des fonctionnalités liées à la cohérence des tables, et dispose de

la fonctionnalité « trigger »[PW12], qui permet de déclencher des actions sur certaines opérations.

MySQL était, quant à lui, plus un « dépôt » de données, au sens où le serveur avait un moteur minimum qui stockait les données dans des tables sans forcément de cohérence entre celles-ci ; on ne pouvait pas parler de base de données relationnelle ; chaque table était indépendante des unes des autres. Ses performances et ses limites sont largement au-dessus de ce nous demandons, car d'après [PW13], des utilisateurs de MySQL utilise celui-ci avec 60 000 tables et 5 000 000 000 (milliards) de lignes... Ce logiciel fonctionne aussi bien sous Windows que sous Linux, et possède plusieurs moteurs de tables. Il a été popularisé avec le développement de technique web, son utilisation facile et simple avec le serveur Apache. Il s'est complexifié et est devenu un véritable SGBD, se dotant des fonctionnalités des autres SGBD exposés précédemment. Il reste très polyvalent du point de vue de montée en puissance tant pour le nombre de requêtes que pour la dimension de la base de données. Il existe de nombreux outils permettant de consulter et de travailler avec MySQL, aussi bien sous Windows et sous Linux.

Au vu de la taille des données, et de mes connaissances dans cette base de données, le choix de MySQL est tout à fait justifié.

IV.9.4 LAMP

A partir des choix exposés ci-avant, le choix pour le serveur web Apache devient incontournable. Ce serveur web est un des serveurs web les plus populaires [PW14], il est utilisé sur différents systèmes.

En conclusion de cette partie IV.9 Choix des plateformes, le choix se porte sur l'ensemble logiciel qui est désigné sous l'acronyme LAMP pour Linux, Apache, MySQL et PHP. De plus, pour Windows, il existe aussi une distribution de logiciel qui à l'exception de Linux, permet d'avoir le serveur web Apache, la base de données MySQL et le langage PHP, Cette distribution facilite l'installation de tous ces programmes sur cette plateforme. Ainsi, l'application pourra être facilement utilisée d'une plateforme vers l'autre.

V Organisation du développement

Cette partie du document décrit l'organisation liée au projet. Ce projet étant en concurrence avec d'autres projets, il a été nécessaire de lui affecter du temps de travail, des ressources de personnes et du développement.

V.1 Organisation administrative

Le projet est encadré par un chercheur de mon laboratoire, directeur de recherche ayant sa propre équipe. De plus, ce projet s'inscrit dans une politique à long terme de valorisation de la recherche au niveau du LAGEP ; mon supérieur hiérarchique, le directeur du laboratoire a accepté de me confier ce projet.

Depuis plusieurs années, j'avais déjà consacré une partie de mon temps de travail à des projets de longue durée: projet de développements, gestion des serveurs ou test de nouveaux logiciels. Afin de mener à bien ce projet, j'ai affecté une priorité haute aux événements ayant trait à celui-ci. Seules les urgences de sécurité, les réquisitions demandées par mon supérieur hiérarchique et les dysfonctionnements liés aux appareils collectifs avaient une plus haute priorité.

Le cadre de travail étant défini, il fallait aussi prendre en compte la dimension humaine du projet.

V.2 Organisation humaine

J'ai été associé à différents travaux de chercheurs durant l'ensemble du projet et Particulier avec Mlle Claudia Elizabeth Mora-Huertas, doctorante qui travaille sur les solubilités et le calcul des paramètres de médicaments.

En tant qu'informaticien, j'ai été mis en parallèle aux travaux de mes collègues

chimistes et du génie des procédés. Ceux-ci jouant le rôle, tour à tour, de mes clients et le rôle d'experts dans la réalisation et la validité des résultats. Travaillant dans le même laboratoire, je dispose d'un accès privilégié. Ainsi, des réunions informelles et à travers des échanges de courriel, j'ai gardé le contact tout le long du projet avec le maximum de souplesse.

L'organisation du développement, sur le plan administratif et humain, étant en place, nous devons maintenant définir les tâches à accomplir.

V.3 Organisation de la réalisation

Cette partie se divise en deux sections. La première décrit la plate-forme : où se situent les serveurs web de publication et d'évaluation, où se situe la base de données, quels logiciels ont été utilisés. La seconde donne une vue générale sur le déroulement du projet.

V.3.1 Plate-forme

Afin de répondre aux différentes contraintes, dont celle de disponibilité, de diffusion, de facilité de maintenance, citées au chapitre trois, la plateforme de développement sera en php5, sur un serveur web Apache, en version 2, avec l'utilisation d'une base de données Mysql, version 5.

Cette plate-forme existe sur deux des serveurs dont j'ai la responsabilité, les deux en debian GNU/Linux. Le premier a comme adresse lagep.cpe.fr, alias lagep.univ-lyon1.fr. Sa visibilité est réduite aux établissements de l'université Claude Bernard et de l'INSA Il a d'autres rôles, mais il héberge le serveur web que je destine à un rôle d'intranet (au sens d'accessible dans le périmètre ci-dessus) et de serveur de test. Le deuxième est public. Sa visibilité est mondiale et porte le nom de www.lagep.cpe.fr, ayant plusieurs alias dont www.lagep.univ-lyon1.fr, et c'est le serveur web du laboratoire.

Le premier serveur sera celui d'évaluation et de mise au point. Ceci permettant que les utilisateurs puissent continuer à utiliser une version n-1 du site sur le deuxième serveur. Dès que je juge la version en développement exploitable ou corrigeant un bug bloquant l'utilisation, elle viendra écraser l'ancienne version de façon transparente.

Le troisième serveur, lageptx19.cpe.fr, visible que dans le domaine de CPE, contient le serveur de base de données, comme l'analyse le précise. Tant que le besoin ne se fait pas sentir, une seule base de données est gérée. Elle est possède le nom de CPSM.

L'outil principal de développement est Adobe CS3, dont Dreamweaver possède des fonctionnalités intéressantes décrites ci-dessous. Par contre, ce logiciel n'est pas toujours stable. Par exemple, l'utilisation des requêtes, le traitement des données ou de leur visualisation a quelques problèmes comme la perte de la liaison requête-code ou certains affichages cachent d'autres parties. Cela rend presque impossible la modification d'attributs.

CS3 est un ensemble de logiciels dont le produit principal, à mes yeux, est la suite du logiciel Dreamweaver. Ce logiciel possède des fonctionnalités dont celles ci-dessous sont importantes :

- Nous avons un gestionnaire de site, qui permet de mettre à jour un site distant via plusieurs protocoles comme FTP ou FTP sécurisé. Il autorise aussi la possibilité d'avoir un serveur d'évaluation, permettant d'avoir un site en exploitation et un site en cours de développement. Nous pouvons visualiser les différents fichiers sur les différents sites. L'application cherche à maintenir le site local et le site d'évaluation le plus synchronisés possible.
- La fenêtre principale permet de visualiser une page web sous trois formes : la première permet de placer et de disposer les différentes zones textuelles ou graphiques en respectant les normes et langages dont le HTML. La seconde visualisation correspond à l'affichage du code, avec un coloriage syntaxique et une assistance à l'écriture du code. La troisième est un mixte entre les deux, divisant la fenêtre de visualisation en deux parties. En haut, le code de la page positionné sur ce

que l'utilisateur peut voir dans la partie en bas. Tout au long du travail sur la page, le logiciel tente de maintenir une cohérence entre le code et ce qu'il affiche, facilitant grandement la vue d'ensemble de l'utilisateur.

- Différentes aides et assistances génèrent et récupèrent des champs d'une base de données. Le logiciel utilise un système de gestion d'accès aux données bien réalisé. En premier lieu, il faut avoir créé une requête, paramétrable avec des variables liées à la page, et au besoin avoir rempli les informations pour accéder à la base de données. L'application exploite immédiatement les informations pour permettre de voir au plus vite les champs et les données associés à cette connexion. Nous avons, par exemple, la liste des champs de la requête. Nous pouvons, alors, soit par glisser-déposer soit grâce à l'utilisation de menus, visualiser des champs ou générer des tableaux dynamiques contenant une partie ou la totalité des enregistrements sélectionnés par la requête.

Les outils associés à la base MySQL ont été utilisés afin de permettre de créer les différents comptes, les tables de la base, l'insertion et la correction de certaines données.

V.4 La gestion du projet

Initialement, le projet a été découpé en phases dont les noms et les valeurs de temps ont été reportés dans le tableau 5, ci-dessous. J'expose ensuite la méthode utilisée pour la gestion de ce projet. Puis, je fais un état des lieux de ce projet au moment de la rédaction de ce mémoire.

V.4.1 Projet initial

Le document, qui a permis de lancer le projet du point de vu du Cnam, donnait le projet en six étapes :

Étapes	Dates	contenu
1) Cahier des charges	Mai	<p>constitution du cahier des charges, réunion avec les utilisateurs et l'équipe encadrante.</p> <p>Maquettage d'un prototype avec un début de base de données.</p>
2) Correctif du cahier des charges	Juin	<p>Affinage du cahier des charges afin de définir l'importance des différentes fonctionnalités et les hiérarchiser par ordre de priorité décroissante.</p> <p>En parallèle maquettage de la structure des données, du site web et transformation en réalisation du logiciel.</p> <p>Etude des importations à partir de logiciels commerciaux ou open source (faisabilité)</p>
3) Application primaire	Juin et Juillet	<p>Réalisation de l'application cœur à partir du cahier des charges correspondant aux fonctionnalités principales ; intégration de la notion de mélange de solvant.</p> <p>Mise en place d'un distinguo entre paramètre calculé et paramètre expérimental.</p>
4) Deuxième Application	Juillet à Octobre	<p>Extension aux macromolécules, mise à jour des formats de saisie et d'importation.</p> <p>Extension aux macromolécules : nécessite d'utiliser d'autres lois dont la théorie est en évolution (sujet de thèse en cours).</p> <p>Prise en compte d'un troisième paramètre comme la température (polymères thermosensibles). L'objectif est de pouvoir évaluer le comportement de paramètre de solubilité suivant la température pour un polymère donné.</p>
5) Bilan et Troisième Application	octobre et décembre	<p>Bilan et modification du cahier des charges afin d'intégrer les manques.</p> <p>Bilan : ce que peut faire le logiciel. Quels sont les calculs manquants, modification des calculs suivant l'évolution de la théorie.</p> <p>Modification du cahier des charges en fonction du bilan et développement des nouvelles fonctionnalités</p>

Étapes	Dates	contenu
6) Phase de validation	Janvier	Validation des calculs par les utilisateurs et les expérimentateurs. Validation de l'ensemble du logiciel avant la mise en disponibilité générale.
	et Février	

Tableau 5: Phase initiale du projet

Ce découpage et les délais proposés n'ont pas été respectés. Certaines étapes ont été mélangées afin de répondre au plus vite à des demandes ou en réaction à des tests. Actuellement, l'application est en étape 3, la « première application ». Celle-ci ne contient pas toutes les fonctionnalités annoncées dans l'étape 3 du tableau 5 ci dessus. Par contre, certaines demandes ont été déplacées ; en effet, le mélange de solvants n'y est plus. Elle a été reportée dans une application ultérieure. Je me suis adapté aux différentes contraintes de temps, de disponibilités des personnes et d'urgence de certaines tâches influençant ce projet. Toutes ces contraintes sont décrites dans le chapitre VII. Contraintes rencontrées page 103.

Je fournis un peu plus loin, un tableau et un diagramme de Gantt sur le déroulement réel du projet. La réorganisation du projet est en relation avec l'utilisation d'une méthode, dont je vous livre le détail dans la section suivante.

V.4.2 Un peu de méthode

J'utilise une méthode de développement en V mais j'y apporte deux modifications principales dues au fait que je sois le seul développeur :

- Durant la phase du codage, je fais des tests et je corrige immédiatement aussi bien au niveau de cette phase qu'au niveau architectural. Les tests unitaires et d'intégration sont souvent fait en même temps et renvoie sur la conception détaillée.
- J'ai pris le parti de faire une maquette, puis une première application et continuer

jusqu'à ce que le produit nous convienne. Cette méthode correspond plus à des techniques semi-itératives, mais sans prendre en compte l'élément temps, qui pour les cycles semi-itératifs est un facteur important.

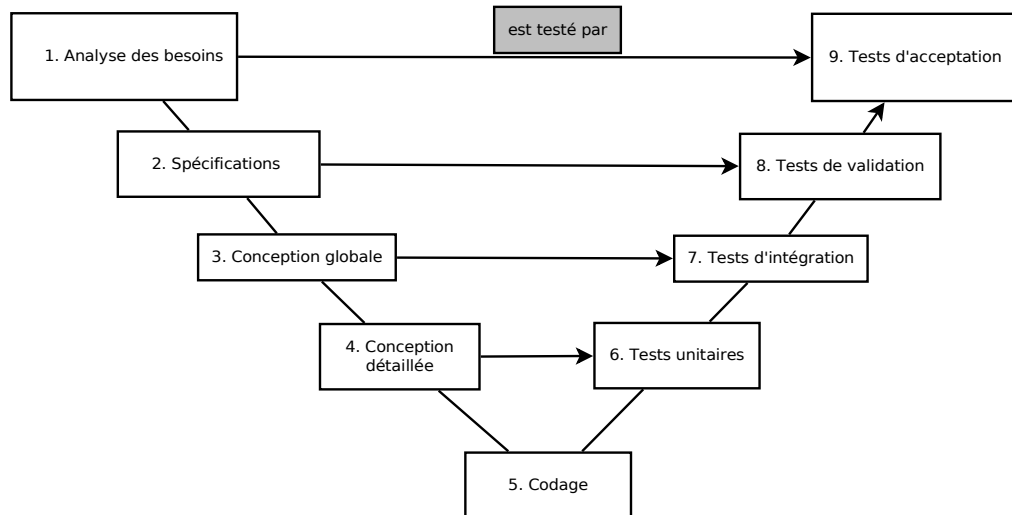


Figure 9: Cycle de vie en V [WP15]

Afin de permettre de faire le lien avec le tableau 6 des phases à la page suivante et les phases en développement en V de la figure 9, je donne les équivalences suivantes :

L' « Analyse des besoins » regroupe les phases du projet ci dessous : l'appréhension du projet, le cahier des charges, la recherche d'outils qui se suivent dans le temps et validation du Projet contenant la faisabilité et l'acceptation du cahier des charges.

La « spécification » comprend le choix de la plateforme et une partie de l'Analyse. En parallèle, je note la phase d'acceptation du projet par le Cnam, réalisée durant cette période.

La « conception globale », correspond à une autre partie de l'analyse. Elle correspond à ce qui est exposé dans le chapitre suivant (VI.Réalisation page 67).

Puis le Maquettage, les essais et tests de la maquette et la première Validation correspondent à la première itération de la « conception détaillée » à « tests unitaires » en passant par le « codage ».

Après le maquetage, nous avons lancé la première application qui reprend en majorité les modules de l'application finale mais qui exclut le traitement lié à la gestion des comptes, et la gestion de sources bibliographiques. Dans la partie dédiée au traitement du chapitre suivant (cf.VI.3. Les traitements page 77), j'exposerai l'étude des éléments dédiés à la gestion de ces deux modules.

V.4.3 Déroulement du projet

Cette section a pour objectif de donner une idée de la réalisation et de son déroulement dans le temps. Je commencerai par un tableau des différentes phases

Nom de la phase	Date de début	Date de fin	Durée (j)
Appréhension du projet	04/05/2009	23/06/2009	35
Cahier des charges	03/06/2009	04/07/2009	23
Recherche d'outils	06/07/2009	01/08/2009	19
Choix Plateformes	03/08/2009	22/08/2009	15
Validation du Projet	24/08/2009	29/08/2009	5
Analyse	31/08/2009	27/10/2009	41
Acception du projet Cnam	10/09/2009	30/09/2009	14
Maquetage	27/10/2009	10/12/2009	31
Essai et test Maquette	10/12/2009	01/01/2010	15
Première Validation	04/01/2010	08/01/2010	4
Correction Analyse	08/01/2010	29/01/2010	15
Première Application Web	29/01/2010	31/03/2010	43
Correction et débogage	31/03/2010	20/05/2010	36

Tableau 6: Phases réalisées du projet

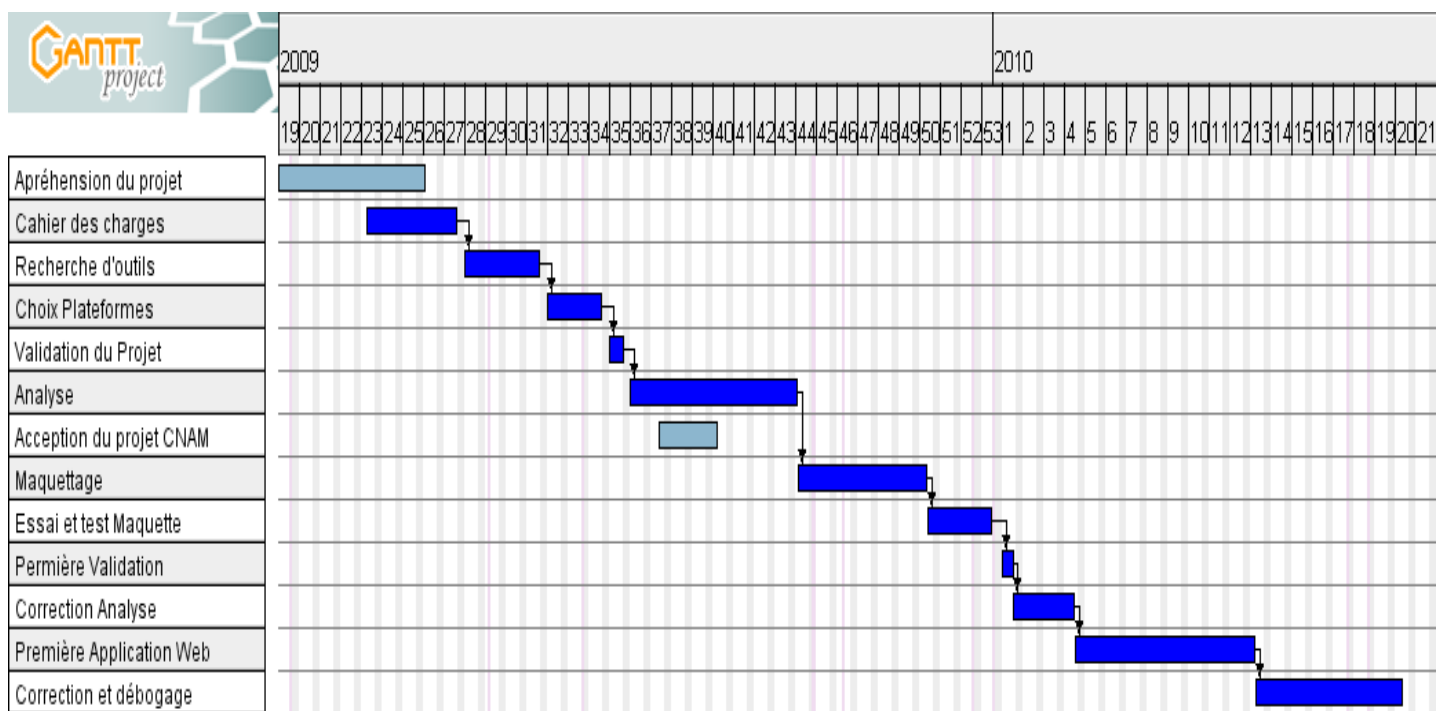


Figure 10: Diagramme de GANTT du projet

Le planning sous forme de diagramme de GANTT du projet est décrit précédemment. Les cycles de développement après les corrections de la première application web n'ont pas été intégrés. Bien que cette phase soit notée jusqu'au 20 mai, elle est prévue jusqu'à la fin juillet. Les unités sont en semaine, le numéro de semaine dans l'année étant en en-tête de colonne, les phases sont données par ligne.

VI Réalisation

Ce chapitre traite de la réalisation de l'application en suivant les choix établis aux chapitres précédents. Ce descriptif prend en compte l'état dans lequel se situe le projet : une première maquette a été réalisée comprenant une base de données incomplète et des pages en PHP. Puis après discussion avec les utilisateurs et mise au point, la phase de la première application web a été lancée. Elle exclut la gestion des sources de documentation et de la gestion des comptes et groupes d'utilisateurs ; considérées comme annexe vis à vis du développement.

Dans un premier temps, je présente la base de données, puis les différentes implémentations des fonctionnalités pré-citées, pour dans un troisième temps donner une idée de fonctionnement de l'application.

Le logiciel est développé sur deux éléments : la base de données et le langage de programmation PHP, comme défini aux chapitres précédents. Le PHP sera utilisé pour implémenter les traitements.

Avant de présenter la base de données, il faut expliquer le pourquoi de certaines solutions de programmation. Ces solutions ont un impact sur la base de données. Sa description viendra ensuite, avec un modèle physique du dernier état de la base de données. Après viendra la description des traitements, suivie du plan du site, et des descriptions des pages du site.

VI.1 Solutions de programmation

Je présente, dans cette partie du document, quelques solutions de programmation pour simplifier le développement de la première application et le maquetage.

VI.1.1 Les champs listes

Certaines relations correspondent à l'idée de « Est dans » une liste, comme par exemple, un utilisateur « est dans » un groupe ou une valeur « est dans » telle liste. Ceci peut être stocké sous forme d'une chaîne de caractères utilisant un formatage spécifique.

Le formatage choisi, à des fins de minimiser le stockage, est le stockage des valeurs souhaitées, délimitées par un séparateur choisi de façon à ce qu'il ne rentre pas en conflit avec les valeurs.

Par exemple, la liste (1 , 2 , 3 , 4) pourra être codée par la chaîne "1;2;3;4", le séparateur choisi le signe du point virgule : « ; ». Le choix aurait pu être « , » si les valeurs stockées n'utilisent pas de notation avec une virgule.

Comme la relation « est dans » peut représenter beaucoup de liens, cette chaîne doit permettre le stockage de l'ensemble des valeurs. Pour une certaine efficacité, il faut permettre que le champ de stockage soit utilisable pour les recherches.

Ce type de solution est assez efficace pour la recherche d'une appartenance. Savoir si une valeur est dans une liste revient à chercher une sous-chaîne dans une chaîne. Il est à noter que la recherche doit être réalisée en bornant les deux chaînes de recherche avec le caractère de séparation.

Par exemple, rechercher le 9 dans la liste "1;5;7;29" doit se traduire par rechercher ";9;" dans ";1;5;7;29;". Si l'utilisateur de ce champs ne le fait pas, il ne cherche que "9", des faux positifs apparaissent, ici 29. Ne mettre qu'un séparateur, avant ou après, ne résout pas le problème complètement : ";9" ou "9;" peut générer d'autre faux positifs, comme dans cette liste "10;95;19;7".

Suivant que le programmeur fait de nombreuses recherches ou pas, il pourra décider de stocker soit la chaîne au plus court, comme "1;5;7;29", soit la chaîne avec les séparateurs, ici, ";1;5;7;29;" pour la même liste. Dans le premier cas, il sera nécessaire de rajouter les deux

délimiteurs de début et de fin avant toute recherche. Dans le second, la recherche pourra être exécutée sans pré-traitement. Dans tous les cas, le programmeur doit finement gérer les délimiteurs de début et de fin de chaîne.

VI.1.2 La table utilisateurs

La table « utilisateurs » regroupe les deux tables « utilisateurs » et « groupes ». Plus exactement, la table est l'union des champs des deux tables. Un champ booléen « EstGroupe » a été ajouté, et deux champs listes, expliqués à la section précédente, ont été ajoutés pour simuler la liste des utilisateurs d'un groupe, ainsi que la liste des utilisateurs qui sont des administrateurs. Bien sûr, pour représenter un utilisateur, ces deux champs sont vides.

VI.1.3 Astuces pour les variables de calculs

Le modèle conceptuel de données détaille la vision d'une solution au niveau des variables de calculs très polyvalentes. La base autorisera l'ajout d'une méthode de calcul, avec sa liste de variables. Chaque couple (molécule, variable) permettra d'obtenir la valeur de la variable pour la molécule. Afin de faciliter le développement des premières applications, j'ai choisi de projeter dans l'entité Molécules, correspondant à la table TMolecules, les valeurs nécessaires au calcul du paramètre de solubilité par la méthode de Van Krevelen.

J'ai ajouté un champ informatif « EtenduMol » qui permettra de faire la liaison, si cette solution devait perdurer, avec le modèle conceptuel de donnée finale. Ce champ, booléen, aura une valeur à Vrai (-1), si des couples (molécule, variable) existe pour d'autres méthodes. Pour celles liées à Van Krevelen, elles sont directement dans la table TMolecules (et par extension Tfragments), et correspondent à V_a , F_{di} , F_{pi} , E_{hi} équivalant aux valeurs $V_{a,i(298)}$, F_{di} , F_{pi} et, E_{hi} données dans les tableaux Tableau 10 en Annexe C: les données page 119.

VI.2 base de données

Je vous présente la base de données dans sa structure à travers la liste des ses tables, dans un premier temps, puis leur description et enfin le modèle physique de données qui existe actuellement.

VI.2.1 Tables

Avant d'avoir la liste des tables et leurs champs, je donne quelques précisions sur les types de données utilisés. La chaîne de longueur limitée correspond au type VARCHAR(255) de MySQL ; la chaîne peut aller jusqu'à deux cent cinquante cinq caractères. Si la longueur est précisée, il est toujours un VARCHAR de la longueur donnée. Le type double permet de stocker une valeur numérique à virgule flottante. Le type entier est une valeur numérique entière même de grande valeur ; elle peut aussi avoir le qualificatif de automatique pour indiquer une incrémentation automatique de la valeur pour la création d'enregistrement. Le type booléen permet de stocker une valeur vraie ou fausse, en MySQL ce type est simulé par une valeur entière TINYINT(1).

La base de données est composée des tables suivantes :

- la table TSegments

Cette table est uniquement présente pour maintenir opérationnelle la maquette, et avoir une table ne contenant que les molécules du référentiel. Elle possède les champs suivants :

- NoFrag, entier, numérotation automatique, clé primaire de la table.
- NomRef : chaîne de longueur limitée, donne le nom de la molécule. Actuellement, comme le champ suivant sans les liaisons de covalence, si elle n'ajoute pas d'information de distinction.
- FormuleTexte : chaîne de longueur limitée, correspondante à la représentation du texte de la formule chimique. Actuellement, aucune mise en forme n'est prévue. Ce champ ne doit pas être nul.
- FormuleImage : chaîne de longueur limitée, correspond à un chemin, relatif,

vers un répertoire contenant les images.

- EtenduFrag : booléen, correspond à la notion d'extension de la molécule dans une autre table.
 - VaFrag : double, paramètre nécessaire au calcul suivant la méthode Van Krevelen. Cette valeur, représentant un volume, peut être négative. Par défaut, ce champ est mis à nul.
 - FdiFrag : double, comme précédent mais représente la valeur de dispersion.
 - FpiFrag : double, stockera la valeur de polarité du fragment.
 - EhiFrag : double, toujours pour les valeurs de Van Krevelen, correspondant aux énergies de covalence.
 - RefSources1 : texte long, contenant la référence du document validant les variables contenues dans le champsSource1.
 - ChampsSource1 : chaîne de longueur limitée et liste de champs précisant pour la portée de la référence indiquée en au champ RefSources1
 - RefSources2 : texte long, pour une deuxième référence
 - ChampsSource2 : chaîne de longueur limitée, pour déterminer l'ensemble de champs ayant un lien avec la référence stockée en RefSources2.
 - Les champs UCree et UMaj, de type entier contiennent les références aux utilisateurs ayant mis à jour ou modifié l'enregistrement. Ces champs ne seront plus détaillés dans les autres tables.
 - Les champs DCree et DMaj, de type date, contiennent la date respectivement de création et de mise à jour. Ces champs ne seront plus détaillés dans les autres tables.
- la table TMolécules :
cette table, correspondant à l'entité Molécule, possède les champs suivants :
 - NoRef : entier automatique, contient l'identification de la molécule et est la clé primaire de la table.
 - NomMol : texte de longueur limitée, il est le nom que donnera l'utilisateur qui le créera.
 - FormuleTexte : chaîne de longueur limitée contenant la formule chimique de la

molécule ou de la structure.

- FormuleImage : chaîne de longueur limitée, correspond à un chemin, relatif, vers un répertoire contenant les images.
 - Formule3D : chaîne de longueur limitée, indiquant le chemin, relatif aussi, vers un répertoire contenant les fichiers décrivant la formule en trois dimensions.
 - PSM : double, stockera la valeur du paramètre de solubilité moléculaire.
 - EtenduMol : booléen, champ indiquant que d'autres valeurs sont stockées dans une autre table ou un ensemble de tables.
 - VaMol : double, paramètre nécessaire au calcul suivant la méthode Van Krevelen. Par défaut, ce champ est mis à nul.
 - FdiMol : double, comme précédent mais représente la valeur de dispersion.
 - FpiMol : double, stockera la valeur de polarité de la molécule.
 - EhiMol : double, toujours pour les valeurs de Van Krevelen, correspondant aux énergies de covalence.
 - Proprio : entier, numéro d'identification de l'utilisateur créateur.
 - DMaj : date.
 - EstFrag : booléen, est un fragment.
- la table TMoleculeElements
Cette table représente la relation « EstSousMolecule ». Sa clé primaire est (NoRefMol, NoRefSSMol)
 - NoRefMol pour l'identification de la molécule
 - NoRefSSMol pour l'association d'une sous-molécule, contenant le numéro unique d'une molécule.
 - NbMol : entier, pour le nombre de sous-molécule.
 - la table Tparams
Cette table ne correspond à aucune entité, ni aucune relation.
 - CodeParam : chaîne de longueur limitée, taille 5, clé primaire, correspond au nom du paramètre.
 - ValParam : chaîne de longueur limitée, contenant la valeur stockée sous forme de

chaîne

- uniteParam : chaîne de longueur limitée, permettant de saisir l'unité de la valeur.
 - DomaineParam : chaîne de longueur limitée, à des fins de regroupement de plusieurs paramètres
 - LibelleLong : chaîne de longueur limitée, permettant de stocker le libellé long. Il est destiné à être affiché.
 - LibelleCourt : chaîne de longueur limitée, 32 caractères au maximum, contenant un texte le plus synthétique possible pour le paramètre.
 - Autrecode : chaîne de longueur limitée, indique une autre variable.
- la table TSaisieMol
- La table TSaisieMol contiendra une copie des informations pour l'édition d'une molécule. Elle reprend la plupart des champs de la table TMolecules :
- SessionID : chaîne de longueur limitée, contenant l'identifiant de la session utilisateur. Elle est la clé primaire de la table.
 - NomMol : chaîne de longueur limitée, contenant le nom de la molécule.
 - FormuleTexte : chaîne de longueur limitée, la formule chimique de la molécule ou de la structure.
 - FormuleImage : chaîne de longueur limitée, correspond à un chemin, relatif, vers un répertoire contenant les images.
 - Formule3D : chaîne de longueur limitée, indiquant le chemin, relatif aussi, vers un répertoire contenant les fichiers décrivant la formule en trois dimensions
- FormuleImage :
- PSM : double, champ de la valeur du calcul du paramètre de solubilité.
 - EtenduMol : booléen, indique que d'autres tables contiennent des informations.
 - ValMol : double, stocke la valeur Va
 - FdiMol : double, contient la valeur Fdi
 - FpiMol : double, stocke la valeur Fpi
 - EhiMol : double, contient la valeur Ehi.
 - Proprio : entier, identifiant de l'utilisateur .
 - DMaj : date, date de mise à jour.

- RefMol : Entier, référence à la table TMolecules. Elle contiendra le numéro de la molécule ou la valeur zéro si l'enregistrement représente une nouvelle molécule.

- la table TUtilisateurs
Représente les entités « utilisateurs » et « groupes ».
 - AdresseMail : chaîne de longueur limitée, clé primaire de la table. Elle contiendra l'adresse courriel du compte.
 - uid : entier automatique, numéro unique de l'utilisateur
 - NomCourt : chaîne de longueur limitée, contenant le nom public de l'utilisateur
 - EstGroupe : booléen, définit si cet « utilisateur » est un groupe
 - LstGroupe : chaîne de longueur limitée, la liste des utilisateurs dans ce groupe.
 - LstAdminGroupe : chaîne de longueur limitée, la liste des administrateurs du groupe.
 - MotDePasse : chaîne de longueur limitée, contenant le mot de passe crypté.

- la table TmpMoleculeElements
Cette table est le pendant à la table TMoleculeElements. La clé primaire est le couple (SessionID,NoRefFrag).
 - Proprio : entier, le numéro du propriétaire.
 - NoRefMol : pour l'identification de la molécule
 - NbMol : entier, pour le nombre de sous-molécule/sous-fragment.
 - SessionID : chaîne de longueur limitée, identifiant de la session

- la table TssFrag
la table TssFrag est la table représentant la relation qui existe entre les fragments et les sous-fragments. Sa clé primaire est (NoFrag,NoSSFrag).
 - NoFrag : entier, numéro unique d'un fragment.
 - NSSFrag : entier, pour le pendant de la relation, le numéro unique du fragment associé au sous-fragment.
 - TagCalcul : chaîne de longueur limitée, désigne à l'origine une relation à la méthode de calcul.

- Pondération : entier correspondant au nombre de sous-fragments.

VI.2.2 Descriptions des tables

TFragments est la table contenant le référentiel, il sert à obtenir des données qui sont "certifiées". Elle correspond à une zone de dépôt, si besoin est de remettre à zéro ou produire une nouvelle base "vierge". Elle est accompagnée par la table TssFrag, modélisant la relation EstSousMolecule (dans ce EstSousFragment).

TMolecules et TMoleculeElement sont les tables cœur de logiciels. TMolecules contient les données liées aux molécules, tant celle du référentiel, marquée comme fragments, que celles créées par les utilisateurs. TMoleculeElements est la table qui gère la relation EstSousMolecule, elle contient les liens entre une molécule et ses sous-molécules

TParams est une table de gestion des paramètres, qui est définie comme étant un n-uplet (code, valeur, type, domaine, libellé long, libellé court, autre code). Elle permet de gérer des valeurs qui devront être stockées et propres à la base et à l'application. Le champ « domaine » contiendra une valeur donnant une idée de la destination de ce paramètre, par exemple pour regrouper des paramètres entre eux. Le champ « uniteParam » permet de préciser l'unité du paramètre.

TSaisieMol et TmpMoleculeElements sont deux tables de données temporaires permettent les modifications d'une molécule par un utilisateur, à travers une session. La molécule est copiée de TMolecules, avec l'ajout d'un attribut session, et la table TmpMoleculeElements contient la relation EstSousMolecule, avec l'ajout de l'attribut session, permettant de faire le lien avec la table TMolecules. Le choix de IdSession comme clé primaire n'autorise qu'une seule édition par navigateur et pour un compte. Ce choix permet de réduire les erreurs et les conflits mais n'est pas définitif.

La table TUtilisateurs gèrera, comme expliqué à la section précédente, l'ensemble des données touchant à la gestion des utilisateurs et des groupes.

VI.2.3 Modèle physique des données

Le modèle physique de donnée a été produit à partir d'une base Microsoft Access 2007, ce qui permet de l'obtenir d'une façon graphique, plus agréable à lire. Il faut préciser que la représentation ne sait pas gérer les liens avec la même table sur le même champ. Nous avons ainsi deux tables T Fragments et TMolecules, qui ne sont présentes que pour permettre d'afficher les relations entre les tables, respectivement, TssFrag-T Fragments et TmoleculeElements-TMolecules.

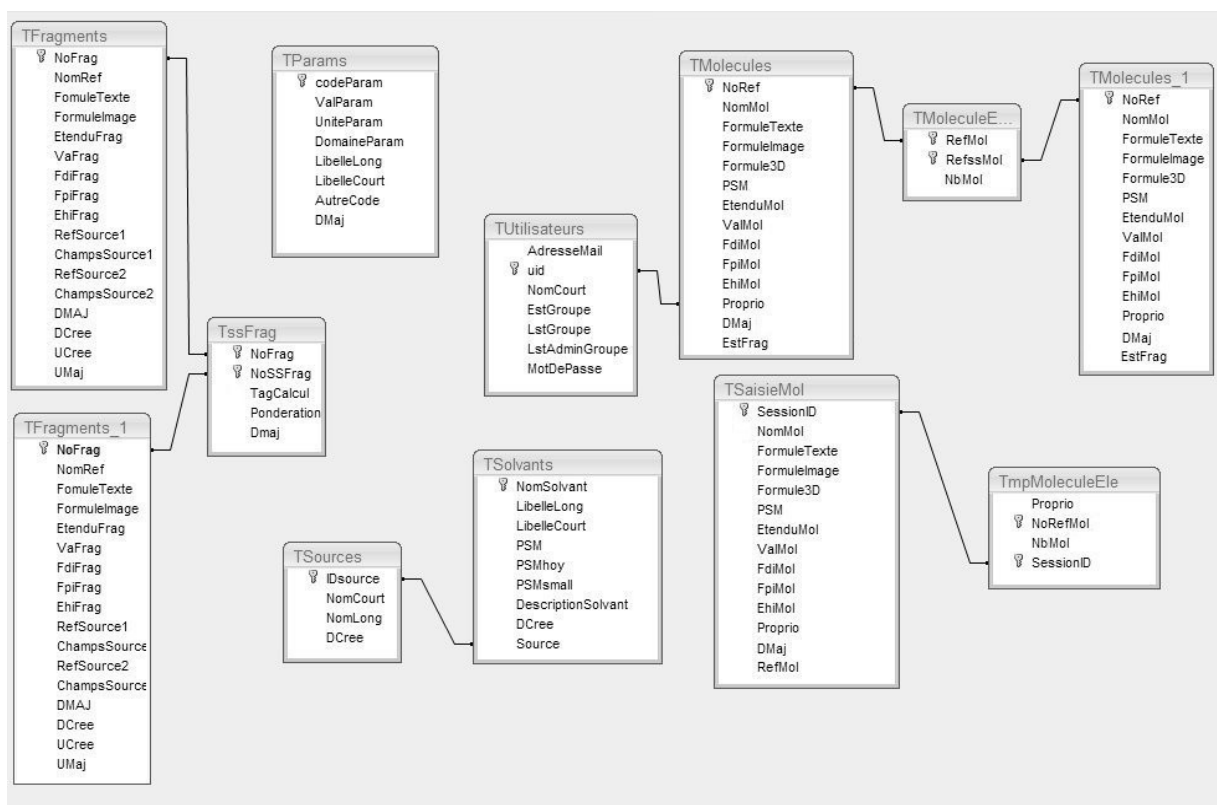


Figure 11: Modèle Physique de données actuel

Il est important de comprendre le fonctionnement des données, qui représente les liens de relations entre les différentes tables.

Après avoir présenté les tables afin d'expliquer le fonctionnement de l'application, je vais présenter les traitements des données, leur organisation à travers le plan du site, ainsi que les traitements effectués au niveau des pages du site web.

VI.3 Les traitements

Je donne ci-dessous l'ensemble des traitements à réaliser avec le détail des procédures employées, qui permettent de faire la liaison entre la partie chimie et la partie programmation. La liste des traitements est donnée sans découpage, mais chacune des lignes est reprise en étant regroupée par thème : gestion utilisateurs, traitement de la construction d'une molécule, et le calcul du paramètre de solubilité.

Les traitements à réaliser sont :

1. ajout d'un nouvel utilisateur.
2. identification d'un utilisateur.
3. changement de mot de passe.
4. demande d'un nouveau mot de passe.
5. demande d'entrer dans un groupe.
6. ajout d'un utilisateur dans un groupe.
7. chargement des tables temporaires.
8. ajout ou modification de sous-molécules.
9. création d'une nouvelle molécule.
10. copie d'une molécule.
11. sauvegarde d'une molécule.
12. gestion des molécules disponibles (de la personne, visible et public).
13. Effacement des tables temporaires.
14. le calcul du paramètre de solubilité.

VI.3.1 Gestion utilisateurs

Cette section regroupe les traitements touchant à la gestion des utilisateurs

VI.3.1.1) Ajout d'un nouvel utilisateur

Ajouter un utilisateur demande plusieurs traitements et plusieurs évaluations. En paramètre, il nous faut un quadruplet (nom, email, mot de passe, confirmation). Le module doit faire comme vérification :

- le mot de passe n'est pas vide.
- le courriel n'est pas vide.
- le nom n'est pas vide.
- le mot de passe et la confirmation sont identiques.
- l'adresse de courriel n'existe pas dans la base.

Si tous ces tests ont été passés avec succès, le module peut insérer la personne dans la base des utilisateurs et confirmer par courrier électronique si l'opération est bien réalisée. L'utilisateur commence l'identification avec les paramètres courriel, et mot de passe.

Remarque : nous n'avons pas de test réalisé pour une attaque en « SQL injection », au vu de l'opération qui est l'ajout d'une ligne dans la base. Les attaques « SQL injection », est l'ajout d'une commande SQL dans un des paramètres. La protection utilisée est la mise entre guillemets ou entre apostrophe des valeurs fournies au moment de la requête ; cela limite l'interprétation d'un code potentiellement injecté.

Le courriel n'est pas non plus testé. Envoyer un courrier électronique peut devenir un test.

VI.3.1.2) Identification d'un utilisateur

Le traitement pour se faire reconnaître du site. Les paramètres sont la paire (courriel et mot de passe). Les vérifications sont :

- le mot de passe n'est pas vide.
- le courriel n'est pas vide.

Si les vérifications sont passées avec succès, une recherche de la personne sur son

courriel est réalisée. Le mot de passe est crypté pour être comparé au mot de passe stocké, crypté lui aussi. Le module charge l'enregistrement afin de propager les informations de connexion, à l'exception du mot de passe.

VI.3.1.3) Changement d'un mot de passe

La demande de changement de mot de passe, l'utilisateur doit fournir le mot de passe actuel, un nouveau mot de passe et sa confirmation. L'utilisateur doit être enregistré. Les vérifications suivantes doivent être faites :

- le mot de passe actuel n'est pas vide.
- le nouveau mot de passe n'est pas vide.
- le nouveau mot de passe et la confirmation sont identiques.

VI.3.1.4) Demande d'un nouveau mot de passe

La demande de nouveau mot de passe se fait par l'utilisation du courriel. Les vérifications sont :

- le courriel est vide.
- le courriel n'existe pas dans la base.

Si les conditions sont respectées, le module génère un mot de passe aléatoire (par un logiciel externe ou tire une suite de caractère) et l'envoi vers l'adresse électronique sera fait.

VI.3.1.5) Demande d'entrée dans un groupe

L'utilisateur identifié peut demander une inscription dans un groupe. Les paramètres sont le groupe et un texte demandant la raison d'entrer dans le groupe. Les tests à effectuer sont les suivants:

- groupe non vide.
- groupe existant.
- possède une liste d'administrateurs.

Si les conditions sont valables, la recherche du groupe permet, s'il existe, de connaître la liste des administrateurs et ainsi d'avoir une liste d'adresses pour envoyer un courrier précisant le nom public, le numéro d'identification et le texte d'explication, avec la raison de la demande.

Remarquez que seul le numéro sera fourni pour éviter de fournir des adresses courriels.

VI.3.1.6) Ajout d'un utilisateur à un groupe

Cette partie de l'application gère la possibilité d'ajouter, si l'utilisateur courant est administrateur du groupe, un utilisateur qui a fait sa demande. Les paramètres d'entrées sont, l'utilisateur enregistré, le numéro du demandeur et le groupe. Les vérifications sont :

- Le groupe n'est pas vide.
- le numéro existe dans la base.
- l'utilisateur est administrateur du groupe.

Si ces conditions sont valides, un test sur la présence dans la liste des utilisateurs du groupe est réalisé. En cas d'échec, l'utilisateur est ajouté en fin de liste.

VI.3.2 Traitement sur la construction des molécules

Cette partie détaille les traitements sur la construction des molécules, nouvelles ou

existantes. Une molécule est chargée dans les tables temporaires pour permettre son édition. Il n'y a pas de gestion de conflit entre les éditions multiples, car seul l'utilisateur propriétaire peut modifier une molécule lui appartenant ; l'hypothèse est qu'il ne peut pas, involontairement, valider sur deux navigateurs la même molécule en même temps.

VI.3.2.1) Chargement des tables temporaires

Pour l'initialisation, le paramètre est soit un numéro de molécule, soit une valeur nulle, pour une nouvelle molécule.

Si la valeur en paramètre est nulle, le traitement crée un enregistrement vide minimum dans la table de « TSaisieMol » avec le numéro de session, une valeur « indéfinie » pour le nom. Aucun enregistrement pour la table TmpMoleculeEle.

Si la valeur passée en paramètre n'est pas nulle, la sélection de la molécule dans la table TMolecules est faite en fonction de la valeur et le traitement copie les valeurs des champs existants dans TsaisieMol correspondant à Tmolécules. Le numéro de session est utilisé comme faisant partie de la clé primaire. Si un échec se produit, c'est que l'utilisateur utilise déjà cette session pour éditer. La table TmpMoleculeElements est chargée des valeurs de TMoleculeElements, toujours avec l'identifiant de session en complément de clé.

La gestion des erreurs durant ce traitement utilise une variable chaîne formatée en HTML pour fournir à la fin du traitement les différentes erreurs accumulées. A charge du programmeur de l'afficher ou de la traiter différemment.

VI.3.2.2) Ajout ou modification de sous-molécules

Cette partie du traitement touche à l'ajout des sous-molécules pour une molécule. En paramètre, un tableau de couple (NoRef, Quantité) où NoRef est le numéro d'identification d'une molécule et Quantité est la quantité demandée ou sa modification (valeur relative). Le

numéro d'identification de la molécule est considéré comme valide.

Si la quantité n'est pas nulle, on recherche dans la table TmpMoleculeEle, si cet sous-molécule est présente dans la table (toujours avec l'identifiant de sessions pour la complété la clef).

Si le cas s'avère vrai, la quantité est ajoutée à la quantité déjà existante. Si la quantité finale est positive, alors la table TmpMoleculeEle est mise à jour. Si la quantité est égale à zéro ou négative, l'enregistrement devra être supprimé. Si aucun enregistrement existe dans la table TmpMolecule, alors l'enregistrement est ajouté à celle-ci, si la quantité est positive.

Remarque : le traitement actuel ne prend pas en compte le cas où une molécule serait sa propre sous-molécule ce que l'utilisateur pourrait générer indirectement au moment de l'ajout d'une sous-molécule. Il faudrait pour éviter cette situation, maintenir une liste de toutes les molécules et sous-molécules utilisées par la molécule, et vérifier à chaque insertion si la nouvelle sous-molécule n'est pas dans cette liste.

VI.3.2.3) Création d'une nouvelle molécule.

Le traitement de création d'une nouvelle molécule, revient à lancer le traitement « chargement des tables temporaires » avec un numéro d'identification de molécule à la valeur nulle.

VI.3.2.4) Copie d'une molécule

La copie d'une molécule est proche de la modification d'une molécule, Il faut donc appeler la procédure de chargement des tables temporaires et mettre à zéro le champ de référence NoRef. Ainsi l'application croit avoir une nouvelle molécule.

VI.3.2.5) Sauvegarde d'une molécule

Elle n'est possible que si le champ NomMol n'est pas vide, ainsi que FormuleTexte. Si le champ RefMol est égal à zéro, la procédure considère que la molécule est nouvelle et donc ajoutera celle-ci à la table des molécules. Sinon, elle modifiera l'enregistrement correspondant dans TMolecule.

Une transaction est lancée. Puis la procédure supprimera les enregistrements, s'ils existent dans la table TMoleculeElements en relation avec la molécule en cours d'édition, par le champ RefMol.

La procédure copie les valeurs de l'enregistrement de TSaisieMol dans leurs équivalents de TMolecules soit par une commande INSERT pour une nouvelle molécule (RefMol égale à zéro), il faudra récupérer le nouveau identifiant de la molécule, soit par une commande UPDATE, pour une molécule déjà existante. Si une erreur se déclenche, la transaction est annulée, l'erreur est retournée.

Une copie des enregistrements de la table TmpMoleculeElements sera faite dans la table TMoleculeElements. Si une erreur survient, la procédure s'arrête avec une erreur, la transaction est annulée.

VI.3.2.6) Gestion des molécules disponibles

Ce traitement n'est pas encore implémenté et n'est pas prévu avant une nouvelle version du logiciel. Ils permet d'obtenir la liste des molécules visibles pour un utilisateur.

La première étape est de connaître les groupes dont l'utilisateur fait partie. Pour cela, une requête sur la table TUtilisateurs en cherchant dans le champ LstGroupe contenant la valeur du numéro unique de l'utilisateur.

La seconde étape est une boucle sur la liste des groupes ainsi obtenus et pour chacun, de rechercher des enregistrements correspondants dans la table « EstVisible » (non encore implémentée).

VI.3.2.7) Effacement des tables temporaires

L'effacement des tables temporaires est important car toute nouvelle molécule hériterait de l'ancienne liste des sous-molécules. Il faut donc effacer dans la table TsaisieMol l'enregistrement correspondant au numéro de session en cours. Si l'intégrité référentielle n'est pas active, il s'agit d'effacer les enregistrements de la table TmpMoleculeElements sur le même critère.

VI.3.3 Calculs du paramètre de solubilité

Le détail du calcul du paramètre de solubilité est donné dans le cas général, correspondant à l'analyse, et le cas échéant à la projection dans la première application suite aux raccourcis de développement sélectionnés. Ainsi la description sera toujours valable pour les applications suivantes tout en collant à la réalisation actuelle.

Pour calculer le paramètre de solubilité, en utilisant le modèle de Van Krevelen, (voir IV.7.1.Calcul par la méthode de Van Krevelen page 43), nous devons calculer δ et les valeurs qui lui sont associés δ_D , δ_P , δ_H , et V_a . Pour rappel :

$$\delta = \sqrt{\delta_D^2 + \delta_P^2 + \delta_H^2}$$

$$\delta_D = \frac{\sum F_{di}}{V_a} \quad \delta_P = \frac{\sqrt{\sum F_{pi}^2}}{V_a} \quad \delta_H = \sqrt{\frac{\sum E_{hi}}{V_a}} \quad V_a = \sum V_{ai}$$

Ces formules montrent qu'il faut sommer d'une certaine façon les sous-molécules. Il faut déterminer ces valeurs pour chaque sous-molécule ; c'est une méthode récursive correspondant au parcours de l'arbre de décomposition. Vu les types de sommes, il n'y a pas de lien facile entre deux paramètres de solubilité. La condition d'arrêt des appels récursifs, nécessaire pour éviter d'appeler indéfiniment la routine, est de ne plus avoir de sous-molécules ou d'avoir les valeurs au niveau de la molécule en cours.

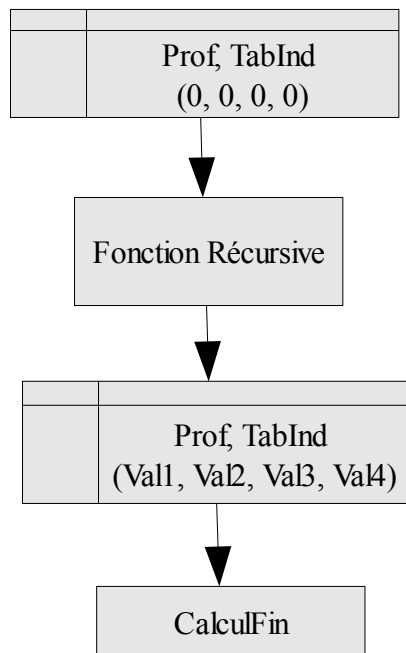
D'un point de vue général, les sommes sont faites sur les variables de la méthode et des opérations sont appliquées après l'obtention de ces sommes. Pour appliquer une méthode, il faut avoir la méthode de somme au niveau de chaque valeur utilisée durant la partie itérative ou récursive et la méthode de calculs après obtention de ces sommes. Ces procédures doivent pouvoir générer un affichage, pour la présentation des calculs ; hors de propos dans cette partie. La procédure de calcul des sommes est notée « Sigma » et celle du calcul de fin est nommé de « CalculFin ». En paramètre, la procédure « Sigma » a besoin, du nom de variable et de la valeur de cette variable, du nombre de fois qu'il faudra sommer cette variable, et d'une valeur correspondant à la somme déjà effectuée.

Un autre problème vient du fait de connaître et de lister les valeurs à sommer. Or seules les méthodes, stockées dans la table « Calculs », définissent quelles sont les valeurs utiles ; il faut construire un tableau d'indicateurs correspondant aux valeurs souhaitées. Si des règles d'addition des variables de même nom diffèrent entre deux méthodes, le tableau aura deux valeurs différentes à maintenir. Ce tableau doit pouvoir stocker des valeurs de type Nulle ou Vide, indiquant que l'information n'a pas été trouvée, ou une valeur numérique à virgule correspondant à la valeur trouvée ; car certaines valeurs ont une valeur de zéro.

La construction du tableau des indicateurs se fait par le questionnement sur la table « Valeurs ». Pour la première application, l'entité n'existe pas en tant que telle, le tableau est donné, son nom est TabInd. Il contient la liste de variables suivantes (Va, Fd, Fp, Eh).

J'utilise une variable dénommée « Prof » pour le calcul de la profondeur. Elle indiquera la distance entre les informations, considérées comme sûres ou validées et le

résultat.

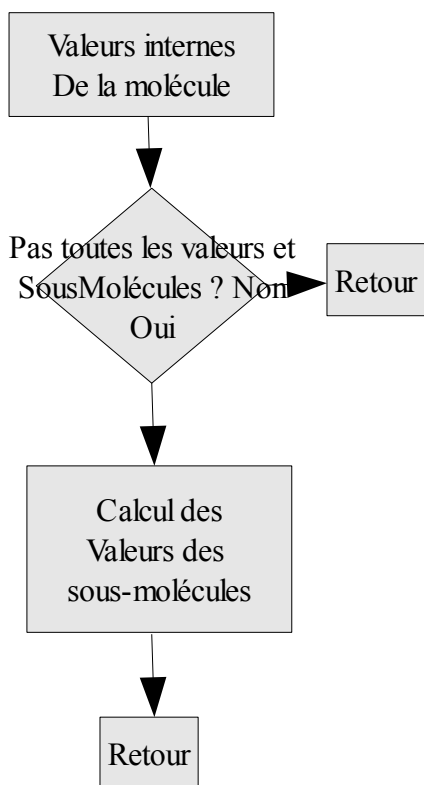


Pour l'initialisation, les valeurs sont mises à zéro. Les calculs finaux ou de sommes doivent prendre en compte le fait qu'une valeur peut être à Null, ce qui invalide le calcul, sur l'ensemble des calculs pour la molécule et de toutes les molécules qui l'utilisent.

Cette procédure appelle le traitement récursif décrit un peu plus bas avec, en référence, le tableau TabInd et la variable Prof (Figure 12 ci-contre)

Le tableau TabInd est fourni en paramètre de la procédure de fin de calcul « CalculFin », qui fera les derniers

Figure 12: Schéma d'appel de calculs. Ces calculs pourront aller jusqu'au bout si toutes les valeurs ne sont pas nulles.



Début du traitement récursif, les valeurs de la molécule sont étudiées. Un tableau, TabIndMasque est construit. Il est le « masque » du tableau des indicateurs : de même dimension, mais contient des variables booléennes. L'objectif de ce tableau est de connaître l'information ajoutée par la molécule en cours de traitement.

Si les valeurs des champs Va, Fdi, Fpi, Ehi sont numériques, le traitement reporte individuellement ces valeurs dans le tableau TabInd, à l'aide de la procédure Sigma. De plus, la variable correspondante du tableau TabIndMasque est mise sur vrai.

Si une seule des valeurs du tableau TabIndMasque est fausse, le traitement va chercher les valeurs manquantes

Figure 13: Grandes étapes du traitement récursif

dans les sous-molécules. Par contre, si toutes les valeurs sont vraies, le calcul s'arrête et retourne à l'appelant pour finir le calcul au niveau au dessus, le cas échéant.

S'il n'y a pas de sous-molécules, Il manque au moins une valeur au calcul. La procédure met à Null les valeurs du tableau TabInd de même indice que les valeurs du tableau TabIndMasque sont à faux, et retourne à l'appelant pour finir le calcul au niveau suivant.

Une copie du tableau TabInd est réalisée. Celle-ci sera utilisée pour le traitement des sous-molécules. En effet, seules les valeurs dont leurs équivalents dans TabIndMasque ont des valeurs booléennes à faux doivent être calculées or cette information n'est pas transmise au traitement récursif.

Pour chaque sous-molécule, le tableau TabInd est transmis par référence à l'appel de la fonction récursive du calcul de paramètre de solubilité. Une copie de variable Prof est fournie à cette fonction, la valeur retournée sera la profondeur nécessaire pour obtenir les valeurs manquantes. La profondeur du calcul sera le maximum des valeurs de profondeurs retournées plus un, correspondant au niveau actuel.

Quand toutes les sous-molécules ont été passées en revue, le tableau TabInd et sa copie sont fusionnés. Le tableau TabIndMasque précise quelle valeur doit être utilisée. Pour chaque valeur fausse du tableau TabIndMasque, la valeur de même indice du TabInd sera remplacée par celle de sa copie.

la procédure retourne à l'appelant, mettant fin à la procédure récursive.

VI.4 Plan du site

Le plan du site permet d'obtenir une vision globale de l'ensemble des pages les unes par rapport aux autres.

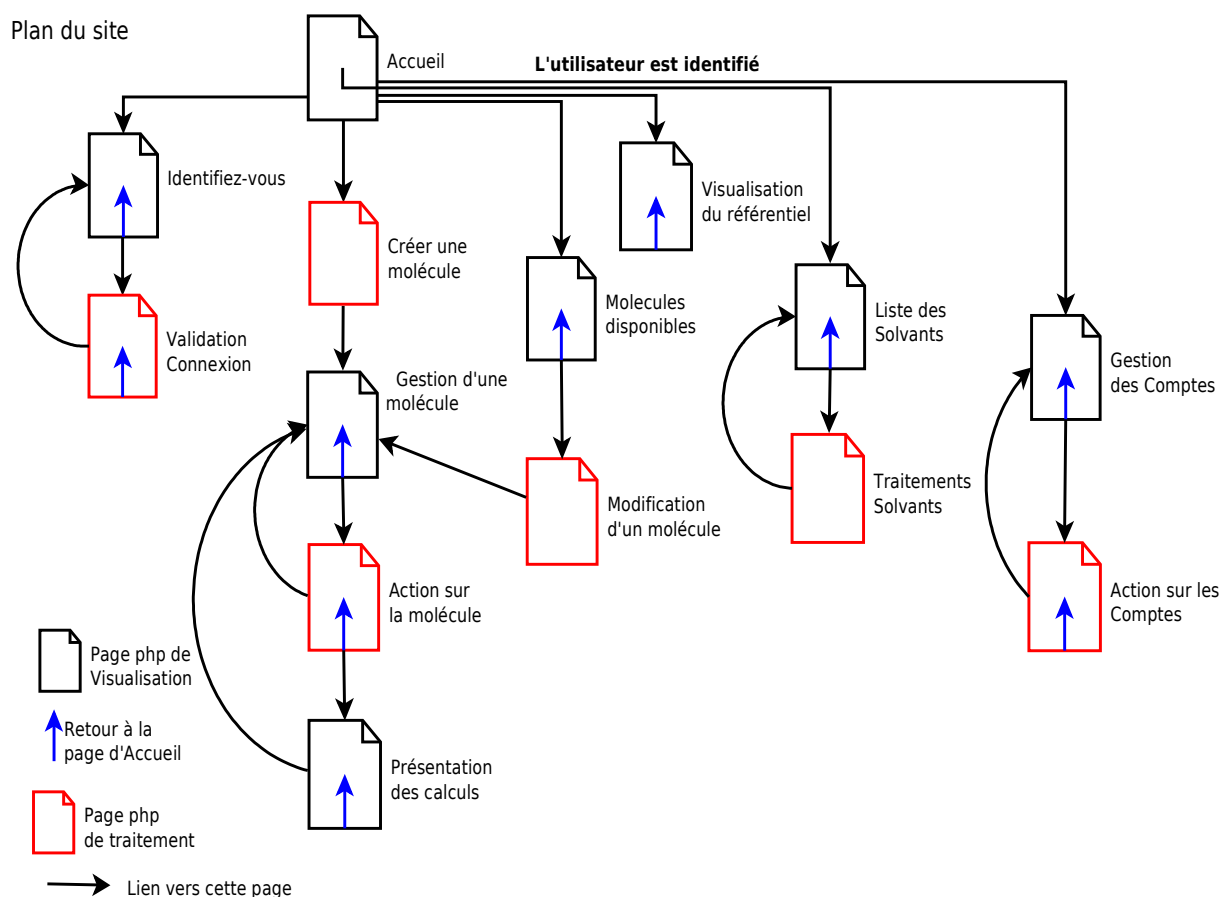


Figure 14: Schéma du plan du site CPSM

La première page du site aiguillera vers les différents modules de l'application. Il est à noter qu'il n'est pas nécessaire de s'enregistrer pour utiliser le site, mais reste indispensable pour sauvegarder ses informations.

VI.4.1 Bandeau des pages

Toutes les pages ont un en-tête reprenant le logo du laboratoire à droite avec un titre à gauche, qui contient des informations supplémentaires ; ici, l'information sur le nom public et son l'adresse courriel qui sert d'utilisateur.



Calcul du Paramètre de Solubilité moléculaire
Molécules disponibles et visibles Pour
OG(garrigues@lagep.univ-lyon1.fr)

Figure 15: Bandeau de page

VI.4.2 Page d'accueil

La page d'accueil permet d'accéder aux différents modules de l'application.



Bienvenue dans le site du
Calcul du Paramètre de Solubilité
moléculaire

vous êtes actuellement **non reconnu**
Pas de mail et autres définitions.

- [Calcul du paramètre de solubilité pour une nouvelle molécule](#)
(fragments et maquette sans sauvegarde)
- [Nouvelle composition de Molécule](#)(en test)
- [Liste des molécules disponibles et visibles](#)
- [Saisie des fragments](#)
- [Bibliothèques des solvants](#)
- Gestion des comptes

[Identifiez vous](#)

Figure 16: Page d'accueil initiale

A partir de la page principale, les choix suivants sont possibles :

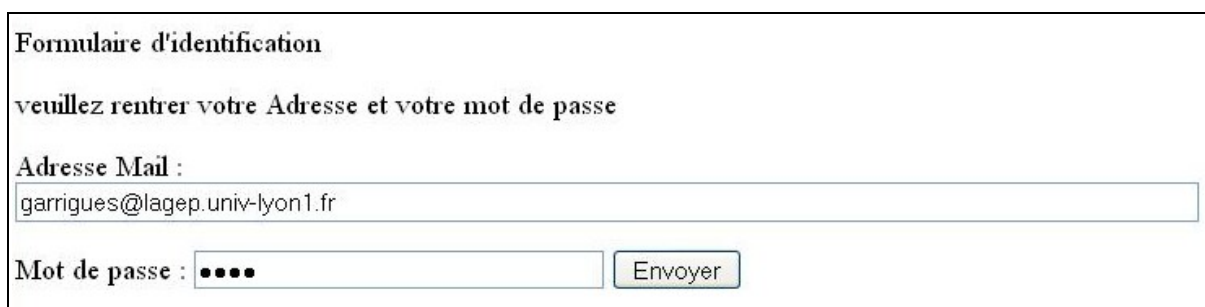
- s'identifier, qui renvoie sur la page « Identifiez vous »
- « Créer une Nouvelle Molécule (fragments et maquette) », elle permet d'accéder pour

l'instant, à l'ancienne maquette, pour permettre de réaliser les tests en parallèle de cohérence.

- « Nouvelle composition de Molecule (en test) », permet la création de composition et envoi sur « Nouvelle Molécule »
- "Liste des molécules disponibles et visibles » afin de modifier une molécule, envoi sur « Molécule disponible », visible par l'utilisateur identifié.
- «Bibliothèques des Solvants » permet d'afficher la liste des solvants grâce à la page « Bibliothèques des solvants »
- « Saisie des fragments » est le lien vers la page « Visualisation du référentiel »
- « Gestion des comptes », visible si la personne est identifiée, amène à « Gestions des comptes ». Dans le cas présent le module gestion des comptes n'est pas encore implémenté. On le voit donc comme n'étant pas un hyperlien.

VI.4.3 Page « Identifiez vous »

Cette page est le point d'entrée pour le module de connexion. Pour s'identifier, l'utilisateur aura la possibilité de :



Formulaire d'identification

veuillez rentrer votre Adresse et votre mot de passe

Adresse Mail :

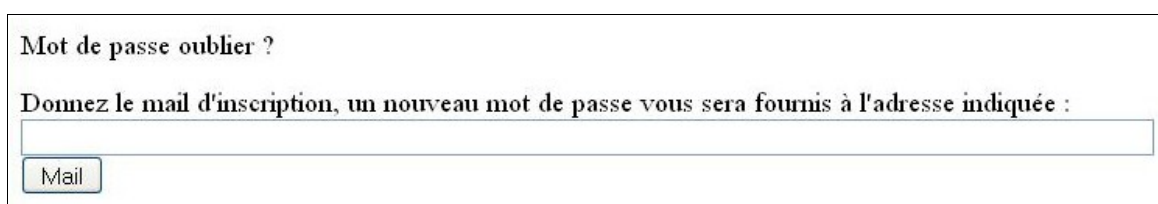
Mot de passe :

Figure 17: Zoom de la page d'identification d'une personne

- saisir une adresse courriel et un mot de passe, puis pourra valider, ce qui passera la main au traitement de la « Validation de connexion »
- créer directement un compte, en saisissant une adresse courriel, un nom public, le mot de passe et sa confirmation. Dans ce cas, le traitement de création de compte sera lancé par la page de «Validation de connexion » après vérification que l'adresse mail

ne soit pas déjà prise.

- pouvoir demander un nouveau mot de passe pour un compte existant. Il y aura envoi d'un courrier contenant le nouveau mot de passe, après vérification de l'existence de l'adresse courriel dans la base. Ce traitement sera réalisé dans la page « Validation de connexion ». L'envoi de courrier ne pourra se faire que si certaines conditions sont réunies, comme la configuration d'un relais de messagerie ou l'accès au réseau. Les serveurs web du LAGEP en disposent, mais ce n'est pas le cas de tous les serveurs web.



Mot de passe oublier ?

Donnez le mail d'inscription, un nouveau mot de passe vous sera fournis à l'adresse indiquée :

Figure 18: Zoom de la page d'identification : mot de passe perdu

- saisir un nouveau mot de passe, après avoir saisi l'ancien mot de passe, puis valider le nouveau. Le traitement sera réalisé dans la page « Validation de connexion ».

Si aucune erreur n'est détectée, la page retourne automatiquement sur la page d'accueil, avec cette nouvelle présentation : en rouge, des informations de débogage. Sinon, le module revient sur la page « identifiez vous » en précisant l'erreur au début de la page.



**Bienvenue dans le site du
Calcul du Paramètre de Solubilité
moléculaire**

vous êtes actuellement **Connecté**

Mail : garrigues@lagep.univ-lyon1.fr UID :2 NomCourt :OG EstGroupe : 0 LStGrp:..

- [Calcul du paramètre de solubilité pour une nouvelle molécule](#) (fragments et maquette sans sauvegarde)
- [Nouvelle composition de Molécule](#) (en test)
- [Liste des molécules disponibles et visibles](#)
- [Saisie des fragments](#)
- [Bibliothèques des solvants](#)
- Gestion des comptes

[Identifiez vous](#)

Figure 19: Page d'accueil avec l'utilisateur identifié

La figure ci dessus montre que la page d'accueil donne certaines informations quand un utilisateur est identifié. Ces informations pourront disparaître dans le futur car elles sont à des fins de débogage.

VI.4.4 Page « Créer une molécule »

Cette page est le point d'entrée pour la création d'une molécule, la gestion et l'interface graphique étant déléguées dans une autre page, « gestion d'une molécule », décrite un peu plus bas.

Cette page aura pour objectif de permettre la création des lignes temporaires dans la base données pour l'édition de la molécule. Elle ira automatiquement vers la page « gestion d'une molécule », sauf si des erreurs sont détectées de façon bloquante : l'impossibilité de créer une ou plusieurs lignes temporaires, les erreurs d'accès à la base ou tout autre problème qui pourrait survenir.

Le nom de la molécule sera positionné sur « indéfinie », aucune sous-molécule ne lui sera affectée. Le numéro de session du client web sera utilisé comme lien durant tous les traitements. Il est aussi la clé primaire des tables temporaires.

VI.4.5 Page « Molécules disponibles »

Dans cette page, l'utilisateur, devra être enregistré et identifié. Il aura la liste des molécules lui appartenant, et celles auquel il a accès. Cette page possède plusieurs onglets qui permettent de sélectionner la vue choisie. J'ai utilisé les fonctionnalités de génération et de gestion automatique des onglets par le logiciel Dreamweaver : il génère automatiquement le code en Javascript pour permettre aussi bien de voir en mode édition qu'en exploitation le rendu visuel.

Plusieurs possibilités lui seront offertes pour les molécules lui appartenant : modification de la molécule, copie d'une nouvelle molécule à partir de celle dans la même ligne et suppression, si cette molécule n'est pas une sous-molécule d'une autre.

Mes Molécules	Les Molécules visibles	Les Molécules Publiques		
Liste des molécules de visibilités générales				
Nom de la molécule (No)	Formule Texte	Image	PSM	Actions
N (-21)	>N-			copier
NO2 (-22)	-NO2			copier
S (-23)	-S-			copier
PO4 (-24)	=PO4			copier

Figure 20: Zoom de la page Molécules disponibles : onglets et listes

En ce qui concerne les molécules ayant le droit d'être vues, l'utilisateur pourra en faire une copie. Pour définir les molécules visibles, la règle est que les molécules affichées sont celles qui sont soit dites « publiques » soit la molécule est mise dans un groupe où l'utilisateur est autorisé.

VI.4.6 Page « Gestion des comptes »

Cette page a pour objectif de faciliter la gestion du compte et de ses appartenances. Si le compte est un administrateur de groupe, il a la possibilité d'accepter, de refuser ou de révoquer l'accès aux groupes qu'il gère.

Un utilisateur peut faire une demande pour être dans un groupe ; la demande sera insérée dans la liste des demandeurs. L'administrateur sera prévenu par un message sur la page d'accueil de l'existence de demande.

Un administrateur de groupe peut insérer un groupe dans la liste des utilisateurs. Cette page n'est pas encore réalisée. Au besoin, je le réalise à la main en modifiant la table sous-jacente TUtilisateurs.

VI.4.7 Page « gestion d'une molécule »

Cette page est la page principale ou le cœur du logiciel. Elle, et sa petite sœur « Action sur une molécule », ont en charge la gestion complète d'une molécule. Vous pouvez sauvegarder celle-ci. Le calcul des paramètres de solubilité sera appelé après ou permettra de lancer directement le calcul, modifier, renommer et copier la molécule sous un autre nom.

Pour donner une vision globale de la page, je vous propose une vue d'ensemble à la page suivante. Cette page est découpée en quatre zones, numérotées de 1 à 4 et affichées dans un cercle.

1) La première permet de définir le nom, la formule chimique textuelle obligatoire. Cette

Nom de la molécule
 Indéfinie
 (Seules les sous-molécule listées-ci dessous sont sauvegardées, faites ajouter en bas avant sinon).
 [Sauver et Quitter] [Sauver et Continuer] [Annuler et quitter]

Formule sous forme de texte : _____

Param Sol Molé connu: _____ (VK Actuel :) [Voi Calcule]

session utile : 6e4330c61c0ac23c280611a26d001ebe

Va	Fdi	Fpi	Ehi
(VK actuel.)	(VK actuel.)	(VK actuel.)	(VK actuel.)

Liste sauvegardé des fragments :

Nombre	Nom	Formule Texte	Formule Image	Valeurs Kleeven (Val, Fdi, Fpi, Ehi)	Actions
				(...)	

Fragments Simples

Fragments Complexes | Molécules Publics | Molécule Perso. | Autres Informations

Listes des fragments simples

No de Ref	Nb.	Nom	Formule Texte	Formule Image	P.S.M
-61		Si(CH3)2	-Si(CH3)2-		
-60		NHCONH	-NHCONH-		
-59		OCONH	-OCONH-		
-58		CONH	-CONH-		
-57		OCOO	-OCOO-		
-56		CO-ar	-CO-ar		
-55		SO2	-SO2-		
...					
-26		6member dbl1	-<< _ >>		
-25		6member non arr	-<_> non arromatique		

Ajouter (ajoute les fragementes ayant un nombre différent de 0)

Figure 21: Vue globale de la page gestion d'une molécule

Il manque à cette partie la gestion des sources. C'est ici qu'un chimiste pourra donner ses informations obtenues à partir d'une expérimentation ou de sa recherche bibliographique.

3) La troisième partie donne la liste des sous-molécules. Les valeurs pour Van krevelen sont indiquées. Ci dessous, l'exemple pour la molécule Polycaprolactone [COO(CH₂)₅] :

Nombre	Nom	Formule Texte	Formule Image	Valeurs Kreeven (Val / Fdi / Fpi / Ehi)	Actions
1	COOgen	-COO-general		(23/ 390 / 490 / 7000)	suppr
5	-CH2	-CH2-		(16.37/ 270 / 0 / 0)	suppr

Figure 22: Liste des sous-molécules (zone 3)

zone possède les trois possibilités d'action vis à vis de la modification ou la création de la molécule. L'utilisateur peut enregistrer son travail, puis quitter ou continuer à modifier la molécule en cours. Il peut aussi quitter sans rien enregistrer. L'action de quitter renvoie l'utilisateur sur la page principale.

2) La seconde zone est la gestion des valeurs propres à la molécule. L'utilisateur peut donner une valeur pour le paramètre de solubilité et des valeurs pour les différents paramètres liés au calcul du paramètre de solubilité suivant la méthode Van

L'utilisateur peut aussi supprimer une ligne, enlevant dans ce cas, la sous-molécule de la composition de la molécule éditée.

4) La quatrième zone est une zone à onglet. Elle liste les molécules suivant des critères simples, comme le distinguo entre fragments complexes, molécules publiques, les molécules de l'utilisateur, et les autres molécules. Ci-dessous un extrait de cette zone, en deux parties, séparées par trois petits point, pour gagner de la place car le bouton ajouter en en bas de la page :

Fragments Complexes		Molécules Publics	Molécule Perso.	Autres Informations	
Listes des fragments					
No de Ref	Nb..	Nom	Formule Texte	Formule Image	P.S.M.
-61		Si(CH3)2	-Si(CH3)2-		
-60		NHCONH	-NHCONH-		
...					
-25		6member non arr	-<_> non aromatique		
Ajouter (ajoute les fragments ayant un nombre différent de 0)					

Figure 23: Zoom de la page : Liste de sous molécules disponibles

VI.4.8 Page « action sur la molécule »

La page action sur la molécule est une page de transition ayant pour objectif de gérer les actions sur la molécule en cours d'édition. Elle est appelée par la page de gestion d'une molécule avec un paramètre indiquant l'action de l'utilisateur.

La première action est l'ajout d'une ou plusieurs sous-molécules. L'utilisateur a choisi les sous-molécules en donnant leur quantité. Le traitement de la page devra vérifier s'il n'existe pas déjà des sous-molécules, et modifier la quantité en fonction dans ce cas. La nouvelle quantité remplace l'ancienne, mettre la valeur à zéro d'une sous-molécule correspond à la suppression de la ligne.

La seconde action permet de sauvegarder la molécule soit comme étape de sauvegarde soit pour quitter proprement la page de gestion de la molécule. La page s'occupe de transférer les données des tables temporaires vers les tables de stockage. Cette étape peut être bloquante si certains paramètres ne sont pas correctes comme le nom de la molécule ou les valeurs saisies ne sont pas du bon type ou que la molécule a été modifiée par une autre personne.

La troisième action est pour annuler et effacer l'ensemble des tables temporaires avant de retourner vers la page index du site, ceci permet une modification.

La quatrième action est le calcul du paramètre de solubilité, ici la page donnera sur le page de « présentation des calculs », décrite un peu plus bas.

La cinquième action est la sauvegarde du calcul du paramètre de solubilité, dans la table temporaire qui stocke les valeurs de la molécule.

VI.4.9 Page « visualisation du référentiel »

La page donne la liste du référentiel, ce qui correspond à la liste de tous les fragments qui ont été fournis à l'origine et tels qu'ils sont dans la base.

VI.4.10 Page « présentation des calculs »

La page expose le calcul du paramètre de solubilité suivant les différentes méthodes. La page est découpée en autant de parties, sous forme de formulaire, que de méthodes disponibles. Pour chaque méthode, on observe un tableau donnant les valeurs des sous-molécules avec la profondeur de leur résultat, puis le calcul des sommes et des valeurs intermédiaires. Enfin, pour chaque valeur exploitable, la possibilité de choisir cette valeur comme paramètre de solubilité est affichée. La page renvoie sur « action sur la molécule » pour réaliser cet ordre.

VI.5 Tests unitaires

Pour vérifier le bon fonctionnement du logiciel, j'ai besoin de retour d'information après l'exécution d'un traitement. je ne connais pas d'outils pour les tests unitaires, et ceux-ci, comme « PHPunit » [PW16], utilisent la programmation orienté objet. Donc, j'ai réalisé mes propres tests. Pour le retour d'information, le principe que j'ai appliqué est simple :

- pour les fenêtres qui ont un affichage, certaines informations de débogage sont clairement affichées et devront être effacées après validation. J'opte ici pour un système d'une variable de session, de type booléen, indiquant si l'application est en mode débogage ou pas.
- pour les fenêtres de traitement, qui nécessitent, en cas de redirection automatique, qu'aucun affichage ne se fasse durant le traitement, j'affiche un message ou une suite de messages quand une erreur se produit, bloquant la redirection. L'affichage comprend un lien de sortie pour continuer à tester vers la page qui aurait dû être affichée.

Des tests unitaires des pages suivantes ont été prévus et ont été soit réalisés, soit à refaire, je les regroupe par domaine d'application. J'utilise des tableaux pour structurer les tests. Les tableaux ci-dessous donnent les grandes lignes. La colonne résultat est vide car les tableaux sont faits pour être réutilisés à chaque fois que nécessaire. En annexe, les tableaux de tests unitaires seront détaillés.

Les tests unitaires sont intégrés au code afin de les refaire à la demande et systématiquement. L'appel de traitement et le passage de paramètres sont simulés par des liens web regroupés dans une page. Pour d'autre cas, le testeur devra faire des saisies.

VI.5.1 Tests unitaires de la maquette

Le plan de test de la maquette est donné ci-dessous à l'image de ce qui vient d'être exposé.

Test	Élément à regarder	Résultat
Saisie de valeurs dans la maquette et lancer le calcul	Les valeurs sont transmises correctement	
Affichage du résultat	La liste des fragments utilisés est affichée Les différents paramètres sont affichés	
Affichage du détail du calcul du paramètre de solubilité	calcul du paramètre de solubilité de $\text{COO}(\text{CH}_2)_5$ correspondant au calcul à la main	
<u>Valeurs limites</u>	Ne doit pas bloquer ni l'utilisateur, l'application.	

Tableau 7: Plan tests unitaires de la maquette

VI.5.2 Tests unitaires pour l'identification

L'objectif de ces tests est de vérifier si l'utilisateur est bien identifié et qu'il s'agit de la bonne personne.

Test	Élément à regarder	Résultat
Saisir un nom d'utilisateur et un mot de passe valide	L'utilisateur doit être reconnu dans la page d'accueil	
Saisir un autre nom d'utilisateur et le mot de passe associé	L'utilisateur doit changer dans la page d'accueil	
Saisir son nom d'utilisateur et un mauvais mot de passe	La page d'accueil doit indiquer que l'utilisateur est non connecté	
Création d'un utilisateur	Saisir son courriel, un mot de passe, le confirmer et donner l'information d'un nom public et valider	
Oubli d'un mot de passe	rappel par mail Génération d'un mot de passe	
Changement d'un mot de passe	Le mot de passe a bien été changé et la personne peut s'identifier	
<u>Valeurs limites</u> Le courriel est vide Courriel très long de 260	Ne doit pas bloquer ni l'utilisateur, l'application. une erreur : « courriel vide, recommencer » une erreur : « courriel trop long, recommencer »	

Test	Élément à regarder	Résultat
courriel sans signe @	une erreur : « courriel incorrect, recommencer »	
Mot de passe vide	une erreur : « mot de passe invalide »	
Mot de passe très long	une erreur : « mot de passe invalide »	

Tableau 8: Plan des tests unitaires pour l'identification

VI.5.3 Tests unitaires de la gestion d'une molécule

Les tests unitaires sont regroupés dans le tableau suivant. Le détail des tests unitaires est reporté dans l'annexe « Annexe D: Tests unitaires. » page 125

Test	Élément à regarder	Résultat
Création d'une nouvelle molécule	Les tables temporaires sont correctement initialisées	
Sauvegarde d'une nouvelle molécule	Sauvegarde du nom, du texte de la formule chimique, les valeurs du paramètre de solubilité et de la valeur du paramètre de solubilité moléculaire (PSM). Blocage si le nom et la formule ne sont pas remplis.	
Sauvegarde des sous-molécules et du nombre d'occurrences	Ajout de sous-molécules et le nombre d'occurrences est respecté	
Modification de la quantité d'une sous-molécule.	Le nombre d'occurrences de la sous-molécule à changer.	
Suppression d'une sous-molécule de la liste des sous-molécules sauvegardées	La ligne de la sous-molécule supprimée n'est plus visible et a été retirée de la table temporaire.	
Affichage du calcul par utilisation du bouton sur voir les calculs, et retour sur la molécule en cours	Le calcul est conforme à l'attente.	
Sauvegarde de la molécule	Les tables ont les bonnes valeurs dans les bons champs. Blocage si le nom et la formule ne sont pas remplis.	
Modification d'une molécule	Affichage de la molécule sélectionnée.	

Tableau 9: Plan des tests unitaires pour la gestion d'une molécule.

VI.5.4 Bilan des tests unitaires

L'objectif des tests unitaires est d'apporter une aide dans la mise au point des fonctions de développement et maintenir une cohérence des résultats. A l'aide de ces tests, j'ai mis rapidement en évidence des comportements erronés.

Par exemple, les tests unitaires de l'identification m'ont permis de comprendre qu'une partie de l'information associée à l'utilisateur était toujours la même quelque soit la personne s'identifiant. Le nom changeait bien en fonction de la saisie, mais la recherche dans la base de données ne se faisait que sur le nom et le mot de passe définie comme valeur de test.

J'ai découvert un problème de sémantique lié au logiciel Dreamweaver. Le logiciel permet de générer facilement une requête pour une page. La requête créée peut être paramétrée ; chaque variable mise en paramètre a deux valeurs (fig. 24, ci-dessous) : l'une pour le test (valeur par défaut), la seconde pour son utilisation dans le programme (valeur d'exécution). Le formulaire de saisie de ces deux valeurs n'est pas clair et souvent j'inverse rôle de ces valeurs. Les tests unitaires réalisés ont permis de repérer et de corriger ce problème de façon systématique.

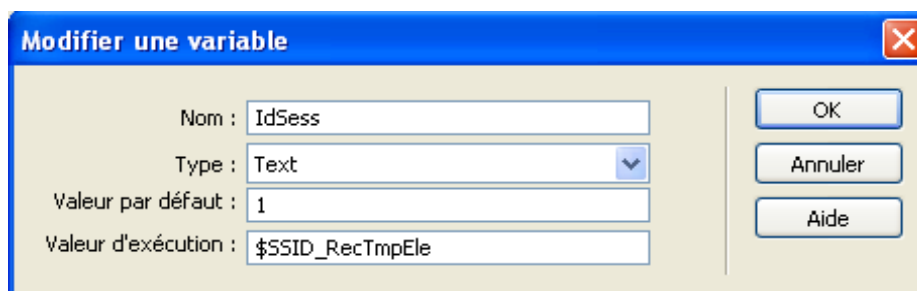


Figure 24: Fenêtre de modification d'un paramètre de requête

Hélas, une partie de ces tests n'a pas pu être automatisée, ce qui ne permet pas de lancer les procédures d'évaluation rapidement ou sans l'intervention d'une personne. Ils demandent beaucoup de temps et leur mise au point revêt un caractère critique ; leur programmation étant faite à la main, est source de nouvelles erreurs.

VII Contraintes rencontrées

Ce chapitre expose les différentes difficultés inhérentes à ce genre de projet ainsi que les événements professionnels marquants qui ont perturbé le projet.

VII.1 Contraintes internes

Les difficultés inhérentes à ce projet ont été principalement liées à la compréhension du fonctionnement du découpage des fragments. Ces informations sont parcellaires et fournies dans différents tableaux. Toutes les valeurs ne sont pas toujours disponibles pour tous les fragments. Une réflexion est réalisée sur la réalité de la notion de fragment : à l'origine, elle était présentée dans notre modèle comme les feuilles de l'arbre de la division moléculaire. Ainsi nous obtenons la garantie d'un résultat si toutes les feuilles appartiennent au référentiel. Or, certains fragments ne possèdent pas toutes les valeurs et doivent être combinés pour être exploitables. Ainsi, nous retrouvons des fragments du tableau 11 parmi ceux proposés dans le tableau 10. Par exemple, nous avons $-\text{CH}(\text{CH}_3)-$ dans le tableau 11 qui est la composition des deux fragments $>\text{CH}-$ et $-\text{CH}_3$ du tableau 10. Le souci vient du fait que $-\text{CH}_3$ ne possède pas de V_a dans nos sources. Pour pouvoir faire les calculs du paramètre de solubilité, il a été nécessaire de modifier le modèle. D'une seule table, nous sommes passés à deux tables : la seconde table représentant la relation entre les fragments afin d'obtenir les valeurs nécessaires au calcul du paramètres suivant la méthode mise en place.

Un second point de difficulté vient de l'utilisation de l'outil, souvent proposée comme une référence dans le développement de site web. Le logiciel Dreamweaver permet de créer des pages web dynamiques rapidement. Vous pouvez voir la même page sous la forme de code HTML et du contenu PHP, ou JavaScript, et en parallèle, la représentation proche du résultat final avec le positionnement graphique des différents éléments. De nombreux outils sont intégrés dont la gestion de site avec la synchronisation avec un serveur final distant et un serveur d'évaluation ou la gestion d'accès à des bases de données.

Le système de gestion d'accès aux bases de données, ainsi que la création de requête SQL associée sont très pratiques à travers une interface graphique. De plus, vous avez la

possibilité de glisser et déposer des champs, de créer un tableau permettant d'afficher une partie ou la totalité d'un résultat de requête de type « Select ». Cela réduit fortement les erreurs de saisies, simplifie le développement et facilite le débogage. Mais la création de requête produisant un code intégré à la page reste sensible à toutes modifications. A de nombreuses reprises, je perdais la liaison avec le code et sa représentation graphique, permettant non seulement le glisser-déposer, mais surtout, la possibilité de modifier la requête pour ajouter des champs ou des critères de sélection. Ce logiciel a aussi des problèmes d'affichage, dans l'exemple ci-dessous, le code nous montre qu'un tableau ayant comme entête « P.S.M », mais son équivalent graphique ne l'affiche pas car il sort du cadre des onglets, qui lui, n'est pas extensible. Si quelque chose doit être modifié, l'utilisateur doit passer par la partie code.

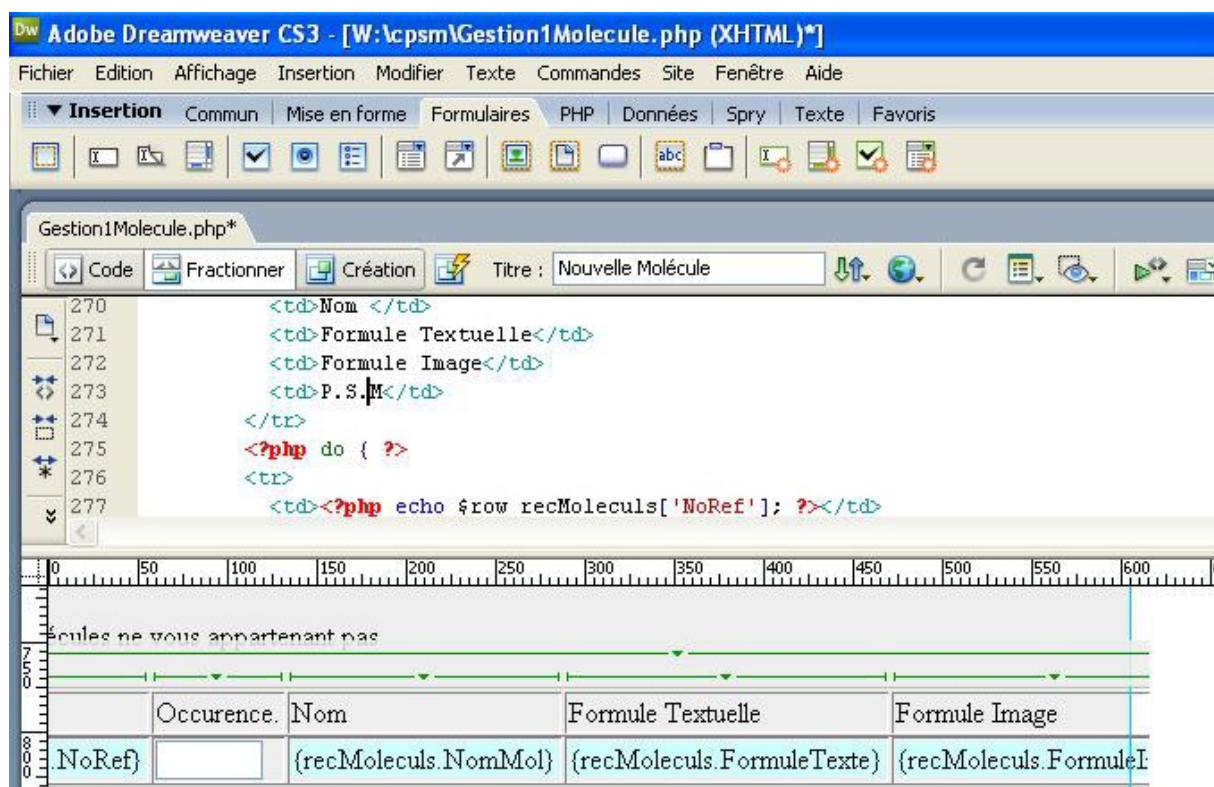


Figure 25: Bug d'affichage de Dreamweaver

VII.2 Contraintes externes

D'autres éléments sont venus perturber le bon déroulement du projet. Des contraintes principalement liées à un environnement de travail changeant. Heureusement, les perturbations ont été ponctuelles mais elles pouvaient durer parfois plusieurs jours de suite. Le laboratoire a déménagé une aile du troisième étage au premier étage incluant, comme je l'ai déjà exprimé avant, la salle informatique et la salle des serveurs. Puis une nouvelle méthode d'évaluation du laboratoire a été demandée, à travers la constitution d'un dossier pour une agence gouvernementale, l'AERES (Agence d'Evaluation de la Recherche et de l'Enseignement Supérieur). En parallèle, au niveau universitaire, une nouvelle structure a été mise en place en fusionnant l'ensemble des UFR de sciences, la Faculté des sciences et technologie. Et pour le CNRS, au niveau des administrateurs systèmes et réseaux, nous avons eu une démarche de mise en place d'une politique de management du système informatique, dont la deuxième phase est actuellement en cours de lancement.

Le déménagement d'une aile complète du LAGEP, a perturbé fortement le fonctionnement du laboratoire. Plusieurs laboratoires d'expérimentations et des bureaux ont été déplacés, certains bureaux ont été modifiés pour recevoir de nouvelles fonctions. Cette aile de bâtiment contenait la salle informatique, mon bureau qui faisait aussi office de salle serveurs et l'atelier de projet d'ingénierie. Afin de réduire au maximum le temps d'arrêt des serveurs et des machines en libre service, une planification rigoureuse a été mise en place. La nouvelle infrastructure, mise au dessus de la structure existante, pour recevoir les nouvelles connectiques aussi bien pour la téléphonie que pour l'informatique, a été testée avant la prise de possession des lieux. Une réflexion sur les besoins en puissance électrique, en climatisation, a été coordonnée entre les différents acteurs. Une nouvelle armoire réseau a été installée dans laquelle l'ensemble de l'aile du premier étage est gérable. L'arrêt et le déplacement, prévus pour une demi journée, n'a réellement pris qu'une petite heure, lié principalement aux multiples câbles et branchements à réaliser pour connecter le tout avec un seul ensemble écran, clavier et souris. La salle informatique a pris plus de temps, car il y avait beaucoup plus de matériel à déplacer et pour chaque machine ou imprimantes, la nécessité de faire la liaison dans l'armoire réseau.

Pour l'évaluation de l'AERES, et à l'opposé de notre évaluation précédente, j'ai été beaucoup plus sollicité. En premier lieu, j'ai concentré toutes les publications stockées dans

différents documents dans d'une base de données. Cette intégration devait numéroter les publications en prenant en compte les différents auteurs, organiser le document final pour que chaque publication soit présente qu'une fois et dans le bon projet. J'ai aussi eu à intervenir sur les documents pour certaines astuces ou apporter des solutions

Du côté de l'université, le président a souhaité réduire le nombre de composantes, et créer une UFR unique de science, regroupant l'ensemble des unités de recherche des sciences des domaines de chimie, biochimie, physique, biologie, génie électrique et procédés, mathématique, informatique et science de la terre. Cette UFR a été désignée sous le nom de « faculté des sciences ». J'ai été élu à son conseil et je me suis proposé comme un de ses représentants dans différents commissions : la commission du personnel, la commission des ressources humaines et le groupe de travail sur les statuts et le règlement intérieur.

De son côté, le CNRS, cherche à renforcer la **politique de sécurité des systèmes d'informations (PSSI)**. Bien qu'ayant une notion forte de politique, ceci étant du ressort du directeur et de conseil de laboratoire, la gestion du système d'information est souvent mis sous la responsabilité de l'informaticien affecté au laboratoire, quand il existe. Comme la tutelle cherche à appliquer cette politique à toutes les unités de recherche attachées à leur établissement, le CNRS a demandé à ce que le directeur nomme un chargé de la **sécurité du système d'information (CSSI)**. Ceci était à l'origine pour relayer les incidents de sécurité liés à l'informatique de l'unité jusqu'aux plus hautes instances. le CSSI a donc été convoqué pour recevoir une formation afin de mettre en place le **système de management du système d'information (SMSI)**, correspondant à la première phase du projet. La deuxième phase est en cours de lancement. Elle demande à ce que le CSSI, avec un groupe de travail, fasse un bilan des actifs, une estimation des risques, une contre proposition pour réduire ou accepter le risque, et suivre une démarche qualité pour vérifier que l'ensemble peut être bien appliqué.

De plus, l'université se restructure à travers le regroupement d'UFR et les services centraux. Ainsi des pôles de gestion et la **direction du système d'information (DSI)** sont apparus. J'ai eu à gérer deux types d'incidences. La première est le passage très rapide et mal contrôlé à une application SIFAC, choisie par l'Agence de **Mutualisation des Universités et**

Établissements (AMUE). La seconde est la fédération à travers un bureau virtuel de la messagerie.

SIFAC a été un logiciel dont le paramétrage fut long et difficile. Cela a généré un ensemble de problèmes. Étant l'informaticien de deux laboratoires, j'avais à analyser et à comprendre à quel niveau se situait le problème : est-ce qu'il y avait une méthode pour contourner le problème ? J'ai donc suivi la formation aux logiciels, étant accessoirement désigné binôme des secrétaires afin de garantir une bonne utilisation dans chaque laboratoire. Une fois formé et le fonctionnement acquis avec l'utilisation de l'environnement associé aux logiciels ; mon aide a été très importante au début de l'utilisation de SIFAC, principalement pour le passage d'un environnement à un autre. Mes collègues secrétaires ayant naturellement un caractère autonome, mon support a été de moins en moins utile.

Enfin, une nouvelle mission a été confiée à la DSI : fédérer l'ensemble des boîtes aux lettres sur une seule forme, aussi bien pour les étudiants, dont le Bureau virtuel de Rhône-Alpes arrive à échéance, que pour les laboratoires. La DSI a cherché à supprimer les serveurs mails installés dans les unités de recherche afin de garantir l'utilisation de leurs boîtes aux lettres. Sans vouloir rentrer dans les détails, qui seraient hors sujet vis à vis du projet, cette obligation, non officiellement déclarée, a nécessité des réunions et des discussions pour comprendre le mode opératoire et nous garantir une continuité de service durant une période assez sensible quant à la circulation du courrier. La migration n'a toujours pas eu lieu par manque de garanties et de temps à affecter à ce projet.

Ces quelques impondérables ont produit différents effets sur le projet, principalement, un décalage dans les temps des objectifs. La partie suivante décrit le bilan du projet et donne les perspectives que j'envisage pour le projet conjointement avec les intervenants principaux.

VIII Bilan et perspectives

Ce projet a été conçu à partir d'une demande très précise d'une équipe de recherche dirigée par le Docteur Hamid Elaissari ; il s'agissait de rendre accessible le calcul du paramètre de la solubilité.

Dans un premier temps, j'ai donc réalisé une analyse complète de leurs besoins en leur proposant la conception d'un logiciel intégrant leurs contraintes. J'ai choisi la plateforme et l'environnement de développement LAMP. A ce stade, nous avons convenu de la planification du travail.

Une maquette a été réalisée et après validation, j'ai développé une première application. L'utilisateur a testé et validé les calculs et la proposition de l'interface. Ce dernier a apprécié la facilité d'utilisation du logiciel. Ce premier test, encourageant, a permis de mettre en perspective certaines difficultés liées à l'arbre de décomposition de molécule, ainsi la nécessité d'améliorer la cohérence des données.

Actuellement la première application est en cours de débogage et d'évaluation par les utilisateurs. Une deuxième application est prévue, elle intégrera les propositions des utilisateurs et les fonctionnalités qui ne sont pas encore implémentées.

A terme, la gestion des utilisateurs pourrait être déléguée à une autre application ; première étape d'une proposition d'un Web-service.

L'un des buts ultimes de ce projet est de réaliser une base de données internationale alimentée par les chercheurs eux-même, après validation de l'administrateur des données scientifiques. Une bonne partie de celle-ci sera réalisée par les élèves de CPE à moyen terme, en leur demandant d'exploiter au maximum les potentialités de ce projet pour le calcul du paramètre de solubilité d'une catégorie de molécules fortement demandée en particulier celui du domaine biomédical.

Une autre ouverture de ce projet permettra de fournir aux utilisateurs ou aux

consultants de ce logiciel, une comparaison directe des paramètres de solubilité calculés par celui-ci et ceux expérimentaux de la base de données.

Un dernier point sera d'extrapoler la limite de ce projet aux calculs des paramètres de solubilité à tous les types de macro-molécules.

En conclusion, ce projet a permis que je m'intègre dans une équipe de recherche et de produire un logiciel qui dépassera, nous l'espérons, largement les frontières du laboratoire.

Le déroulement de ce projet n'a pas été sans heurt. D'une part, j'ai dû assimiler des informations loin de mon domaine ayant arrêté la chimie après le baccalauréat. D'autre part, j'ai eu à affronter une pression de plus en plus grande avec l'augmentation du nombre de personnes à gérer et l'évolution à tous les niveaux de mon contexte de travail.

Néanmoins, le dynamisme de l'équipe de recherche fut un moteur important pour la réalisation du projet ce qui a permis une avancée satisfaisante à ce jour. Les différents défis et les difficultés ont été aplanies grâce aux aides, aux explications et aux différents soutiens de mes collègues, mes enseignants et de mes amis.

Je n'aurais pu arriver à ce résultat sans les méthodes de travail acquises au Cnam et dans ma vie professionnelle. Cette expérience enrichissante m'a rendu plus efficace dans mon travail en me donnant des armes pour mettre en place des stratégies et des méthodes afin de réaliser de futurs projets de cette envergure.

Annexes

IX Annexe A : Bibliographie

IX.1 Sites internet

[PW1] <http://www.universalis.fr/encyclopedie/solvants/>. Vu en 2010.

[PW2] <http://www.cnrs.fr/Cnrspresse/n13a3.html>. Les nanocapsules, un nouveau système de vectorisation et de délivrance des molécules actives. Vu en 2010.

[PW3] <http://cirs.univ-lyon1.fr/reseau/indexplan.html>. CISR, plan du réseau. Vu mai 2010.

[PW4] <http://www.chemistry-software.com/modelling/13052.htm>. Molecular Modeling Pro. Vu en 2010.

[PW5] <http://www.atinternet-institute.com/fr-fr/equipement-internaute/systemes-d-exploitation-janvier-2010/index-1-1-7-190.html>. Document sur la part de visite des systèmes d'exploitations. Source AT internet 2010.

[PW6] <http://www.zdnet.fr/actualites/informatique/0,39040745,39382893,00.htm>. Part de marché des serveurs suivant les système d'exploitations, source IDC, article de Zdnet. 2008

[PW7] <http://www.gnu.org/philosophy/free-sw.fr.html>, le logiciel libre. Vu en Mai 2010.

[PW8] <http://www.sqlite.org/> page d'accueil de SQLite, consulté en octobre 2009

[PW9] <http://msdn.microsoft.com/fr-fr/express/aa718378.aspx>. SQL Server 2008 Express Edition. Vu en octobre 2009.

[PW10] <http://www.oracle.com/technology/products/database/xe/index.html> Oracle Database 10g Express Edition. Vu octobre 2009

[PW11] <http://www.postgresql.org/>, page d'accueil de PostgreSQL, consulté en octobre 2009

[PW12] <http://www.postgresql.org/docs/8.1/interactive/triggers.html>, trigger sous postgresql, consulté en novembre 2009

[PW13] <http://dev.mysql.com/doc/refman/5.0/fr/features.html>. Les fonctionnalités de Mysql. Vu en 2009

[PW14] <http://www.journaldunet.com/solutions/dossiers/chiffres/technoswebserveurs.shtml>

Part de marché de Apache, consulté en mai 2010.

[PW15] <http://perso.ec-lyon.fr/bertrand.david/gl-cnam/GL-CNAM-Intro.PDF>. Diapo 17, B.

David, vu Mai 2010.

[PW16] <http://www.thermodynamique.com/spip.php?article6>. Entalpie. Vu en Mai 2010.

[PW17] <http://www.gnu.org/>. Le projet GNU. Vu en Mai 2010

[PW18] <http://www.idc.fr/about/about.jsp>, IDC France, page « au sujet ». Vu en Mai 2010.

[PW19] <http://www.php.net/manual/fr/preface.php>. Page de documentation PHP. Vu en Mai 2010.

IX.2 Livres

[HJ49] Hildebrand J.H., Scott R.L. « The Solubility of Non-Electrolytes », Reinhold, New York, 3rd ed., 1949.

[VK76] Van Krevelen D.W., Hoftyer P.J., « Properties of Polymers: their estimation and correlation with chemical structure », 2nd ed. Elsevier Scientific Pub, Co., chap 7, (1976)

[VK09] Van Krevelen D.W., Te Nijenhuis, K., 2009. Properties of polymers. Their correlation with chemical structure; their numerical estimation and prediction from additive group contributions, Elsevier, Amsterdam.

IX.3 Autres support

[MH11] Mora Huertas C. E., en cours (2^{ème} année), "*Elaboration de nanocapsules contenant un actif pour usage pédiatrique*". Les directeurs sont le professeur Hatem Fessi et Dr. Hamid Elaissari.

[FG07] Farhad Gharagheizi, « QSPR Studies for Solubility Parameter by Means of Genetic Algorithm-Based Multivariate Linear Regression and Generalized Regression Neural Network ». *10.1002/qsar.200630159*. 22 Janvier 2007.

[RJ94] REVILLA Javier, 1994, « Synthèse et caractérisation de tensioactifs polylémérissables porteurs d'un groupement saccharidique. Etude de la (co)polymerisation radicalaire en solution et en émulsion en vue de la préparation de supports hydrophiles pour la fixation de biomolécules », p 44 à 48, CHABERT B., CANDAU F., RIESS G., UNAM (Mexique).

X Annexe B : Glossaire

Access	Logiciel de la suite bureautique Office de Microsoft. Il peut être le client de plusieurs base de données.
AERES	l'Agence d'Evaluation de la Recherche et de l'Enseignement Supérieur
AMUE	l'Agence de Mutualisation des Universités et Établissements
AMPERE	Laboratoire du génie électrique, électromagnétisme, automatique, microbiologie environnementale et application.
AMPLIVIA	Réseau informatique régional haut-débits.
API	Application Programming Interface.
ATM	Asynchronous Transfer Mode, méthode de transfert haut-débit.
C.P.S.M	Calcul du Paramètre de Solubilité Moléculaire
CISR	Centre Inter-établissement pour les Service Réseaux
CNRS	Centre National de la Recherche Scientifique
CPE	deux possibilités : l'école CPE, voir ESCPE ou la Commission Paritaire d'Établissement, organe de l'université pour permettre la promotion ou le suivi de carrière du personnel technique de celle-ci.
CRH	Commission des Ressources Humaines, une commission demandée par le Conseil d'administration de l'université pour débattre de la politique des personnels.
CSSI	Chargé de la Sécurité des Système Informatique
Debian	distribution linux ayant orienté serveur et un système de gestion des paquets logiciels permettant de connaître le degré d'utilisation des logiciels non libres. Ainsi l'administrateur peut choisir d'être pure Opensource ou plus laxiste dans ses choix.
DHCP	Dynamic Host Configuration Protocol, protocole réseau pour faciliter la configuration réseau IP d'une machine.
DSI	Direction du Système d'Information, le service central de la gestion de l'informatique au niveau de l'université.
ECL	École Centrale de Lyon.
EDI	Environnement de Développement Intégré d'un logiciel pour aidé

Entalpie	L'enthalpie est la somme de l'énergie, qu'elle soit sous forme thermique ou mécanique, gagnée ou perdue par un corps [PW16].
ESCOPE	École Supérieure d'Électronique Chimie et Physique.
Fortran	langage de programmation procédurale.
FST	Faculté des Sciences et Technologies
FTP	File Transfert Protocole, protocole réseau pour transférer des fichiers
Gallénique	ancien nom de la pharmacotechnie. Principe de produire des galets, et donc permettre de transporter des substances dans le corps.
Génies des procédés	domaine de recherche étudiant les méthodes de savoir faire à fin d'industrialiser des procédés.
GNU	GNU is Not Unix, projet pour développer un système d'exploitation complet de style Unix, qui soit un logiciel libre [PW17]
HTML	HyperText Mark-up Language
HTTP	HyperText Transfer Protocol, protocole réseau permettant de transférer des pages HTML
HTTPS	HyperText Transfer Protocol Secured, protocole réseau de transfert sécurisé de pages HTML
IDC	IDC est un acteur majeur de la Recherche, du Conseil et de l'Évènementiel sur les marchés des Technologies de l'Information, des Télécommunications et des Technologies Grand Public. IDC est une filiale de la société IDG, leader mondial du marché de l'information dédiée aux technologies de l'information. [PW18]
INSA	l'Institut Nationale des Sciences Appliquées
IP	Internet Protocol.
Labview	langage et environnement de développement intégré pour programmation en langage G
LAGEP	Laboratoire d'automatique et du génie des procédés
LAMP	Linux, Apache, MySQL, PHP ; un ensemble cohérent de logiciels libres afin d'obtenir un serveur web opérationnel.
LIMA	un monomère amphiphile, désigné soit par lipomalonamide ou soit par 11-(N-p-vinylbenzyl) amidoundécanoyl maltobionamide. Utilisé



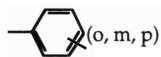
	dans la thèse [RJ94] et donné comme exemple.
Logiciel libre	« Logiciel libre » fait référence à la liberté pour les utilisateurs d'exécuter, de copier, de distribuer, d'étudier, de modifier et d'améliorer le logiciel [PW7]
Lyres	Lyon Recherche Enseignement Supérieur , le réseau métropolitain de l'agglomération lyonnaise
Matlab	logiciel de la société Mathworks, orienté calculs et utilisation de matrices.
MySQL	Système de gestion de base de donnée, logiciel libre
Nanocapsule	capsule de la taille inférieure au micromètre, exprimé en nanomètre.
OpenBSD	système d'exploitation en OpenSource
OpenSource	autre terme pour logiciel libre, essayant de mettre en avant la partie disponibilité du code.
Pare-feu	logiciel ou matériel permettant de protéger un ordinateur ou un réseau d'ordinateurs d'attaques informatique à l'aide de règles
pharmacotechnie	technique liée à la vectorisation d'un médicament ou d'une substance. La vectorisation est le principe de transporter une substance vers une cible.
PHP	un acronyme récursif « PHP: Hypertext Preprocessor », langage de programmation fortement utilisé pour le web. un langage de script HTML, exécuté côté serveur [PW19]
Postgresql	moteur de base de donnée en logiciel libre.
Produit Actif	substance ayant un rôle ou une action.
PSM	Paramètre de Solubilité Moléculaire , nom du projet.
PSSI	Politique de la Sécurité du Système d'Information , choix de la direction d'un laboratoire pour donner des règles de fonctionnement en cas de perturbations du système d'information.
RAM	Random Access Memory , mémoire volatile d'un ordinateur.
Renater	le Réseau National pour la Technologie, l'Enseignement supérieur et la Recherche
RMU	Réseau Métropolitain Universitaire
ROCAD	Réseau Optique du CAmpus de la Doua

SGBD	Systeme de gestion d'une base de données correspondant à un ensemble de logiciels pour la gestion et le stockage de données informatisées.
SIFAC	Systeme d'Information, Financier Analytique et Comptable, une sélection de module faite par l'AMUE du logiciel SAP.
SMSI	Systeme de management du systeme d'information.
Solvant	produit permettant de rendre liquide un autre produit.
SQL Server	solution SGBD de Microsoft
SQLite	Moteur de base de donnée léger.
SSH	Secure S Hell, protocole réseau de sécurisation de connexion à distante et un ensemble de logiciels utilisant se protocole.
UCBL1	Université Claude B ernard L yon 1.
UFR	Unité de F ormation et de R echerche.
UMR	Unité M ixte de R echerche.
Virtualisation	la capacité de simuler plusieurs machines logiques ou plusieurs systèmes d'exploitation sur une machine physique.

XI Annexe C: les données

Les tableaux de éléments chimiques, données initiales. Ces groupes structurels sont les fragments qui seront au minimum distribués en tant que référentiel.


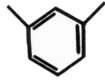
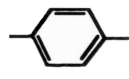
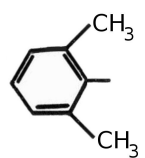
le tableau 7.10 du livre [VK09] page 215 :

Groupe Structurel	F _{di} (MJ/m ³) ^{1/2} . mol ⁻¹	F _{pi} (MJ/m ³) ^{1/2} . mol ⁻¹	E _{hi} J/mol
-CH ₃	420	0	0
-CH ₂ -	270	0	0
>CH-	80	0	0
>C<	-70	0	0
=CH ₂	400	0	0
=CH-	200	0	0
=C<	70	0	0
	1620	0	0
	1430	110	0
 (o, m, p)	1270	110	0
-F	(220)	-	-
-Cl	450	550	0
-Br	550	-	-
-CN	430	1100	2500
-OH	210	500	20000
-O-	100	400	3000
-COH	470	800	4500
-CO-	290	770	2000
-COOH	530	420	10000
-COO-	390	490	7000
HCOO-	530	-	-
-NH ₂	280	-	8400
-NH-	160	210	3100
>N-	20	800	5000

Groupe Structurel	Fdi (MJ/m ³) ^{1/2} . mol ⁻¹	Fpi (MJ/m ³) ^{1/2} . mol ⁻¹	Ehi J/mol
-NO2	500	1070	1500
-S-	440	-	-
=PO	740	1890	13000
Ring	190	-	-
• Symétrie sur un plan	-	0,5 x	-
• Deux plan de symétrie	-	0,25 x	-
• Plusieurs plan de symétrie	-	0 x	0x

Tableau 10: Contribution des groupes de composants aux paramètres de solubilité (Van Krevelen)

le tableau 4.10 du livre [VK09] page 88 :

Groups (Bivalent)	Vai(298) cm3/mol	Vc,i(298) cm3/mol	Groups	Vai(298) cm3/mol	Vc,i(298) cm3/mol
-CH ₂ -	16,37	14,68	-SO ₂ -	(32,5)	(29)
-CH(CH ₃)-	32,72	29,35	-CO-al	13,5	12,25
-CH(t-C ₄ H ₉)-	65,44	58,69	-CO-ar	18,5	16,8
-CH(C ₆ H ₁₁)-	101,76	91,27	-COO-general	23	21,5
-CH(C ₆ H ₅)-	84,18	75,48	-COO-acrylics	20,5	18,4
-CH(p- C ₆ H ₄ CH ₃)-	102,1	91,55	-OCOO-	31	27
-CH(OH)-	22,3	20	-CONH-	(21)	(26)
-CH(OCH ₃)-	40,8	36,6	-NHCONH-	(29)	(26)
-CH(OCOCH ₃)-	52,4	47	-Si(CH ₃) ₂ -	67,5	60,6
-CH(COOCH ₃)-	56,85	51		86	(77)
-CH(CN)-	30,7	27,5		69	62
-CHF-	20,0	18		65,5	59
-CHCl-	30	27,3			
-C(CH ₃) ₂ -	49,0	44,0		104	94
-C(CH ₃)(COOCH ₃)-	100,5	90			
-C(CH ₃)(COOCH ₃)-	74,7	67			
-CF ₂ -	23,7	21			
-CFCl-	33,6	30			
-CCl ₂ -	40,1	36			
-CH=CH-	27,0	24,3			

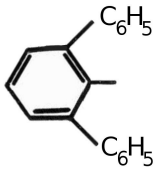

Groups (Bivalent)	Vai(298) cm ³ /mol	Vc,i(298) cm ³ /mol	Groups	Vai(298) cm ³ /mol	Vc,i(298) cm ³ /mol
-Ch=C(CH ₃)- <i>cis</i>	43	40			
-Ch=C(CH ₃)- <i>trans</i>	43	37			
-CH=CCl-	41	37		(208)	190
-C ≡ C -	25	23			
-O- al	(8,5)	(7,9)			
-O- ar	(8,0)	(7,1)			
-NH-	(6,4)	(5,7)		(112)	100
-S-	17,3	15,5			
-S-S-	36	32,5			

Tableau 11: Valeurs recommandées pour les volumes molaires à 298 K

XII Annexe D: Tests unitaires

Chaque plan de test est découpé en deux parties, un rappel du tableau donné dans le texte principal et une suite de tableaux décrivant les détails des tests unitaires. Dans ce deuxième type de tableau, les opérations peuvent être effectuées par l'utilisateur, ils sont en caractère en police normale justifiée à droite ou être une étape du traitement, dans ce cas, la police de caractères est en italique justifiée à gauche. La dernière colonne, notée R, est pour indiquer le résultat. Tout test qui ne se passe pas bien doit être refait.

XII.1 La maquette

Le plan de test de la maquette est donnée ci-dessous à l'image de ce qui vient d'être exposé.

Test	Élément à regarder	R
Saisir des valeurs dans la maquette et lancer le calcul	Les valeurs sont transmises correctement.	
Affichage du résultat	La liste des fragments utilisés est affichée. Les différents paramètres sont affichés.	
Affichage du détail du calcul du paramètre de solubilité	calcul du paramètre de solubilité de $\text{COO}(\text{CH}_2)_5$ correspondant au calcul à la main.	
<u>Valeurs limites</u>	Ne doit pas bloquer ni l'utilisateur, ni l'application.	

1. Saisie de valeurs dans la maquette et lancer le calcul

Opération	Test	R
Mettre 1 dans le premier champ à saisir, puis 2 dans le deuxième, etc...		
Bouton calculer	Avoir la liste des valeurs transmise et ces valeurs correspondent à la saisie réalisée.	
	Les différents calculs sont affichés sans erreur.	

Tableau 12: Test de saisie du nombre de fragments

2. Affichage du résultat

Opération	Test	R
Saisir trois zones texte au hasard dans les nombres occurrences	Les valeurs sont clairement affichées.	
Bouton calculer	La liste de trois fragments est correctement affichée.	
	Le calcul semble cohérent.	
	Les unités des valeurs sont affichées ou indiquées.	

Tableau 13: Test sur l'affichage du calcul

3. Affichage du détail du calcul du paramètre de solubilité

Opération	Test	R
Saisir 1 dans le nombre d'occurrence de COO- Saisir 5 dans le nombre d'occurrence de CH ₂	Les valeurs saisies sont clairement affichées.	
Bouton calculer	La liste des deux fragments est affichée.	
	Affichage du détail du calcul du paramètre de solubilité de COO(CH ₂) ₅ correspondant au calcul à la main.	

Tableau 14: Test avec un exemple COO(CH₂)₅**XII.2 Tests unitaires de l'identification**

L'objectif de ces tests est de vérifier si l'utilisateur est bien identifié et que c'est la bonne personne.

Test	Élément à regarder	R
Saisir un nom d'utilisateur et un mot de passe valide	L'utilisateur doit être reconnu dans la page d'accueil.	
Saisir un autre nom d'utilisateur et le mot de passe associé	L'utilisateur doit changer dans la page d'accueil.	
Saisir son nom d'utilisateur et un mauvais mot de passe	La page d'accueil doit indiquer que l'utilisateur est non connecté.	
Création d'un utilisateur	Saisir son courriel, un mot de passe, le confirmer et donner l'information d'un nom public et valider.	
Oubli d'un mot de passe	Rappel par mail Génération d'un mot de passe.	

Test	Élément à regarder	R
Changement d'un mot de passe	Le mot de passe a bien été changé et la personne peut s'identifier.	
<u>Valeurs limites</u> Un courriel vide Un courriel très long de 260 Un courriel sans signe @ Mot de passe vide Mot de passe très long	Ne doit pas bloquer ni l'utilisateur, l'application. une erreur : « courriel vide, recommencer » une erreur : « courriel trop long, recommencer » une erreur : « courriel pas bien formé, recommencer » une erreur : « mot de passe invalide » une erreur : « mot de passe invalide ».	

1. Saisir son nom d'utilisateur et son mot de passe : L'utilisateur doit être reconnu dans la page d'accueil

Opération	Test	R
Saisie du courriel et mot de passe	Le courriel est en clair et le mot de passe est masqué.	
Bouton envoyer	Le courriel et le mot de passe sont bien envoyés à la page de traitement.	
	Les bons noms de variables sont utilisés comme paramètres.	
<i>Exécution de la requête</i>	Le texte de la requête possède bien les bonnes valeurs.	
	La requête est bien exécutée.	
	Le chargement des variables de sessions est bien fait.	

Tableau 15: Test pour identification d'un utilisateur

2. Saisir un autre nom d'utilisateur et un mot de passe : l'utilisateur doit changer dans la page d'accueil

Opération	Test	R
Saisie un autre courriel et un mot de passe		
Bouton envoyer	La texte de la requête à exécuter est conforme.	
	Le chargement des variables de sessions est bien réalisé.	

Tableau 16: Test de changement d'utilisateur

3. Saisir son nom d'utilisateur et un mauvais mot de passe : la page d'accueil doit indiquer que l'utilisateur est non connecté

Opération	Test	R
Saisie nom et un mauvais mot de passe		
Bouton envoyer	Le texte de la requête possède bien les bonnes valeurs.	
<i>Exécution de la requête</i>	La requête retourne une valeur, le test du mot de passe est faux.	
	Les valeurs de sessions sont à vide, l'utilisateur est déconnecté.	

Tableau 17: Test sur mauvais mot de passe

4. Création d'un utilisateur : saisir son courriel, un mot de passe, le confirmer et donner l'information d'un nom public et valider

Opération	Test	R
Saisie des valeurs	Le courriel, le nom public est en clair et le mot de passe et sa confirmation sont masqués.	
Bouton créer	Les valeurs sont bien envoyées à la page de traitement.	
	Les bons noms de variables sont utilisés comme paramètres.	
<i>Test de doublons</i>	Le texte de la requête est correct et la procédure passe en erreur si celle-ci retourne une valeur.	
<i>Requête d'insertion</i>	Le texte de la requête possède bien les bonnes valeurs.	
	La requête est bien exécutée.	
	Un message de confirmation est envoyé.	
	Les valeurs de session sont chargées avec les bonnes valeurs.	
<u>Valeurs limites</u>	Ne doit pas bloquer ni l'utilisateur, l'application.	
Courriel vide	une erreur : « courriel vide, recommencer »	
Courriel très long de 260	une erreur : « courriel trop long, recommencer »	
Courriel sans signe @	une erreur : « courriel pas bien formé, recommencer »	
Mot de passe vide	une erreur : « mot de passe invalide »	
Mot de passe très long	une erreur : « mot de passe invalide ».	

Tableau 18: Test de la création d'un utilisateur

5. Oubli d'un mot de passe : rappel par mail Génération d'un mot de passe

Opération	Test	R
Saisie du courriel	Le courriel est visible à l'écran.	
Bouton mail	Le courriel est bien envoyé à la page de traitement.	
	Un nouveau mot de passe est généré.	
<i>Requête de mise à jour</i>	Le texte de la requête possède les bonnes valeurs.	
	Le mail est bien envoyé.	
	La requête est exécutée sans erreur. Le mot de passe est changé.	
<u>Valeurs limites</u>	Ne doit pas bloquer ni l'utilisateur, l'application.	
Courriel vide	une erreur : « courriel vide, recommencer »	
Courriel très long de 260	une erreur : « courriel trop long, recommencer »	
Courriel sans signe @	une erreur : « courriel pas bien formé, recommencer »	

Tableau 19: Test sur l'envoi du mot de passe

6. Changement d'un mot de passe : le mot de passe a bien été changé et la personne peut s'identifier

Opération	Test	R
Saisie du courriel, de l'ancien mot de passe, du nouveau mot de pas	Le courriel est en clair et les mots de passe sont masqués.	
Bouton changer	Les valeurs sont bien envoyées à la page de traitement.	
	Les bons noms de variables sont utilisées comme paramètres.	
<i>La requête de mise à jour</i>	Le texte de la requête possède bien les bonnes valeurs.	
	La requête est bien exécutée. Le mot de passe est bien changé dans la base.	
	Le chargement des variables de session est bien fait.	
<u>Valeurs limites</u>	Ne doit pas bloquer ni l'utilisateur, l'application.	
Courriel vide	une erreur : « courriel vide, recommencer »	
Courriel très long de 260	une erreur : « courriel trop long, recommencer »	
Courriel sans signe @	une erreur : « courriel pas bien formé, recommencer »	
Mot de passe vide	une erreur : « mot de passe invalide »	
Mot de passe très long	une erreur : « mot de passe invalide ».	

Tableau 20: Test de changement du mot de passe

XII.3 Gestion Molécule

Plan de tests unitaires pour la gestion des molécules. Rappel du tableau :

Test	Élément à regarder	R
Création d'une nouvelle molécule	Les tables temporaires sont correctement initialisées.	
Sauvegarde d'une nouvelle molécule	Sauvegarde du nom, du texte de la formule chimique, les valeurs du paramètre de solubilité et de la valeur du paramètre de solubilité moléculaire (PSM). Blocage si le nom et la formule ne sont pas remplis.	
Sauvegarde des sous-molécules et du nombre d'occurrences.	Ajout de sous-molécules et le nombre d'occurrences est respecté.	
Modification de la quantité d'une sous-molécule.	Le nombre d'occurrences de la sous-molécule a changé.	
Suppression d'une sous-molécule de la liste des sous-molécules sauvegardées	La ligne de la sous-molécule supprimée n'est plus visible et a été supprimée de la table temporaire.	
Affichage du calcul par utilisation du bouton sur voir les calculs, et retour sur la molécule en cours	Le calcul est conforme à l'attente.	
Sauvegarde de la molécule	Les tables ont les bonnes valeurs dans les bons champs. Blocage si le nom et la formule ne sont pas remplis.	
Modification d'une molécule	Affichage de la molécule sélectionnée.	
<u>Valeurs limites</u> Les nombres trop grand. Nom d'une molécule vide Formule chimique vide	Interdiction par message avant de continuer Refus de sauvegarder avec un message. Refus de sauvegarder avec un message.	

1. Création d'une nouvelle molécule

Opération	Test	R
Cliquer bouton « sauver et continuer »	Les champs Nom et formule chimique sont remplis.	
	La requête d'insertion est bien formée.	
<i>Exécution de la requête</i>	La mise à jour est bien faite.	
	les champs correctement affichés.	
Recommencer une nouvelle molécule		
Cliquer bouton « sauver et quitter »	Les champs Nom et formule chimique sont remplis.	
	La requête d'insertion est bien formée.	
<i>Exécution de la requête</i>	La mise à jour est bien faite.	
	L'utilisateur est sur la page d'accueil.	

Tableau 21: Tests unitaires de création d'une nouvelle molécule

2. Sauvegarde d'une nouvelle molécule

Le test doit être fait en tant qu'utilisateur non connecté : un message doit indiquer que « l'utilisateur n'étant pas connecté, pas de sauvegarde possible. »

Puis le test doit être fait en tant qu'utilisateur connecté.

Opération	Test	R
Cliquer bouton « sauver et continuer »	Les champs Nom et formule chimique sont remplis.	
	La requête d'insertion est bien formée.	
<i>Exécution de la requête</i>	La mise à jour est bien faite.	
	les champs correctement affichés.	
Changer le champ Formule chimique	La requête d'insertion est bien formée	
Cliquer bouton « sauver et quitter »		
<i>Exécution de la requête</i>	La mise à jour est bien faite.	
	L'utilisateur est sur la page d'accueil	

Tableau 22: Test unitaires de sauvegarde d'une molécule

3. Sauvegarde des sous-molécules et du nombre d'occurrence

Opération	Test	R
Chaque zone de saisie dans la zone des fragments et molécule, aura une valeur en commençant par 1 puis 2, etc...		
Bouton « ajouter »	Affichage d'une liste des champs avec la valeur correspondante à celle saisie.	
	Affichage de la requête d'insertion.	
<i>Exécution de la requête</i>	La table TmpMoleculeEle contient les enregistrements.	
Cliquer sur le lien de retour	Affichage de la molécule et des sous-molécules.	

Tableau 23: Tests unitaires pour la sauvegarde des sous-molécules

4. Modification du nombre d'occurrence de sous-molécules.

Opération	Test	R
Mettre « -1 » dans la zone de saisie correspondante au trois première sous-molécules		
Bouton « ajouter »	La première était à 1 doit être supprimée. La seconde passe à 1, et la troisième à 2.	

Tableau 24: Tests unitaires pour la modification de sous-molécules

5. Suppression d'une sous-molécule de la liste des sous-molécules sauvegardées

Opération	Test	R
Utilisation du lien suppr au bout d'une ligne de sous-molécule.	Affichage du code SQL de suppression.	
<i>Exécution de la requête</i>	La ligne a disparu dans la table temporaire.	
Cliquer sur le lien de retour	La sous-molécule disparaît.	
Mettre en négatif le nombre d'occurrence dans la zone de saisie	La sous-molécule disparaît.	
Cliquer sur ajout		

6. Affichage du calcul par utilisation du bouton sur voir les calculs, et retour sur la molécule en cours

Opération	Test	R
Faire une molécule, ajouter des sous-molécules		
Cliquer sur « voir calculs »	Affichage de la page des calculs.	
	La liste de fragments est correctement affichée.	
	Le calcul semble cohérent.	
	Les unités des valeurs sont affichées ou indiquées.	
	Liens de retour (indexe et molécule en cours) sont présents.	

7. Sauvegarde de la molécule

Opération	Test	R
Faire une molécule, ajouter les sous-molécules		
Faire le test unitaire Tableau 22: Test unitaires de sauvegarde d'une molécule page 126	Tous les tests sont passés.	

8. Modification d'une molécule

Actuellement, la page « Liste des molécules disponibles et visibles » n'est fonctionnelle, l'utilisateur pourra utiliser un lien préprogrammé.

Pour un utilisateur connecté, il faut faire les étapes suivantes :

Opération	Test	R
Passer par la page « liste des molécules disponibles et visibles »		
Choisir une molécule et cliquer sur « modifier »	La molécule s'affiche.	
Modifier son nom et cliquer sur « sauver et quitter».		
Revenir sur la page « Liste des molécules disponibles et visibles »	La molécule est modifiée.	

Refaire les opération du tableau ci-dessus pour chaque cas ci-dessous

Opération	Test	R
Modifier sa formule chimique	La molécule est modifiée.	
Modifier les Fpi, Fdi, Ehi et Va	La molécule est modifiée.	

XIII Annexe E : Organigramme hiérarchique

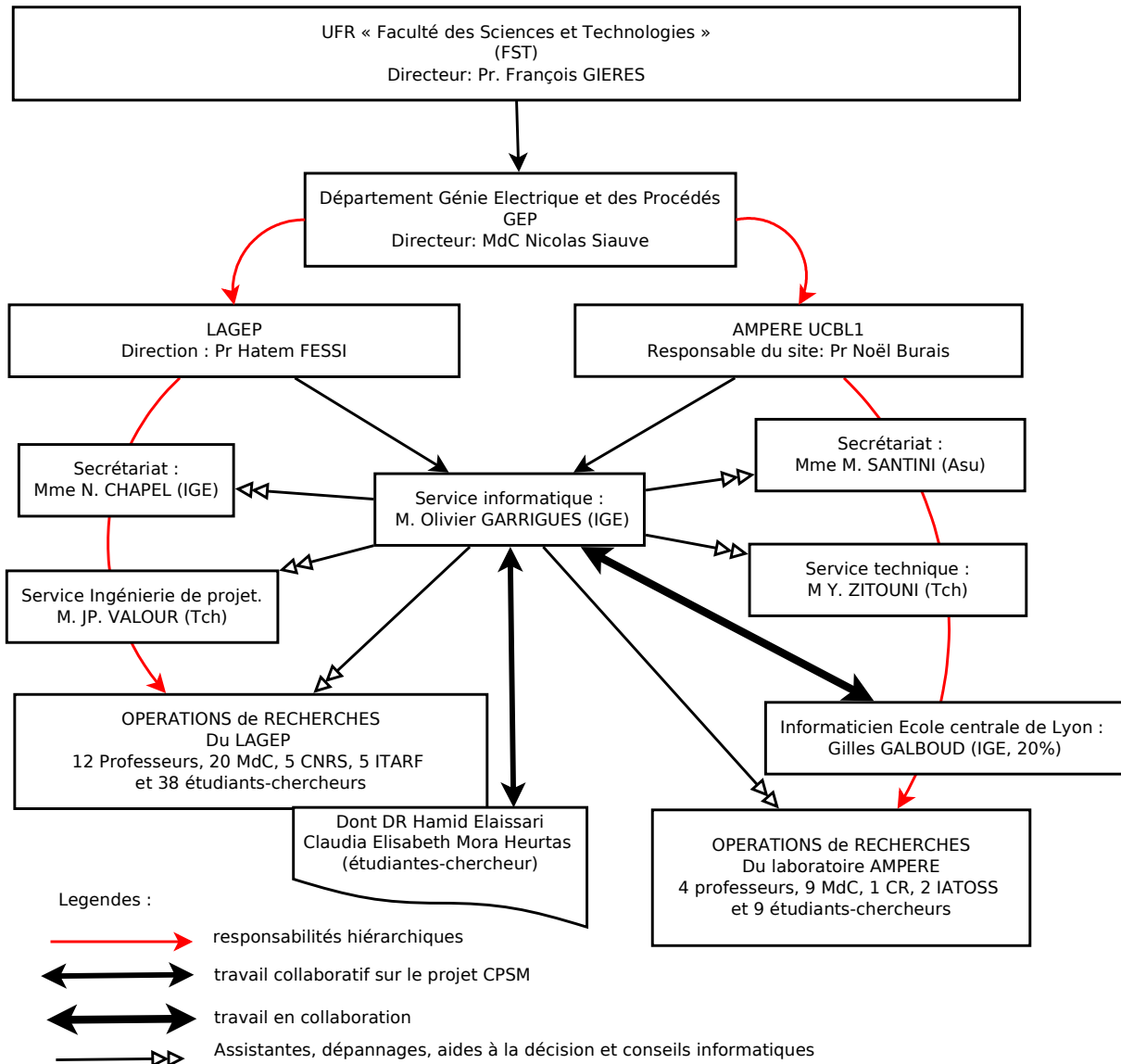


Figure 26: Organigramme hiérarchique au niveau des laboratoires

Résumé et mots clefs

Calcul du paramètre de solubilité moléculaire : projet C.P.S.M. Mémoire d'ingénieur C.N.A.M.
Lyon 2010. Olivier Garrigues.

Résumé

Le LAGEP est un laboratoire dont deux de ses domaines est le génie des procédés et la pharmacotechnie. Les chercheurs du laboratoire utilisent de nouvelles molécules qu'ils cherchent à combiner à des solvants afin de produire de nouvelles substances comme des médicaments. Ayant peu de ces molécules, ils veulent prédire quel sera le meilleur solvant qui permettra de dissoudre leur produit. En évaluant le paramètre de solubilité de chacun des produits, ils peuvent obtenir un début de réponse.

Le projet CPSM a pour objectif de produire une application afin de permettre le calcul de ce paramètre à partir de la décomposition de la molécule en fragments et ce suivant certains critères de fonctionnement, d'utilisation et de méthode.

L'étude proposée mettra en lumière l'analyse qui a été élaborée pour effectuer une approche de solution. De même, les mécanismes mis en œuvre seront détaillés.

Mot clefs : CPSM, solubilité, paramètre, molécule, décomposition

Abstract

The LAGEP is academic laboratory focusing on two main research fields ; chemical engineering and the pharmaceutical technology. The laboratory's researchers are using new molecules in order to formulate new chemical composition for *in vivo* therapy, which needs to investigate the solvency of any new drug in a given solvent. These new molecules are in few quantity, researchers need to predict what will be the best solvent for these molecules before any formulation investigation. The solubility parameter study gives a good preliminary response.

The project, CPSM, aims to construct a software to allow the calculation of the solubility parameters of any molecule using molecular fragments according to performance criteria, method of calculation and use criteria.

The proposed study will pointed out the analysis that has been developed for reaching solution. Similarly, the mechanisms will be implemented in details.

Keywords: CPSM, solubility, parameter, molecule, drug