

G. FLAMENBAUM

E. ABILLON

J. P. BENZÉCRI

**Typologie et analyse factorielle d'un ensemble
de conformations moléculaires : application
à l'Angiotensine II**

Les cahiers de l'analyse des données, tome 4, n° 3 (1979),
p. 339-355

http://www.numdam.org/item?id=CAD_1979__4_3_339_0

© Les cahiers de l'analyse des données, Dunod, 1979, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

TYPOLOGIE ET ANALYSE FACTORIELLE
D'UN ENSEMBLE DE CONFORMATIONS MOLÉCULAIRES :
APPLICATION À L'ANGIOTENSINE II
[CONFORMATION]

par G. Flamenbaum ⁽¹⁾, E. Abillon ⁽²⁾
et J. P. Benzécri ⁽³⁾

Il est commun de recourir à l'analyse statistique multidimensionnelle pour répartir en classes, ou disposer suivant des axes, un ensemble d'objets de l'espace tridimensionnel qui relèvent tous d'un même schéma de structure, mais différent par leurs dimensions et leurs proportions (cf § 1). Au même titre que la crâniométrie des vertébrés, l'étude des conformations d'une molécule peut bénéficier de telles analyses; pourvu que les données soient convenablement codées (§ 2). C'est ce que l'on a fait pour une hormone polypeptidique d'un grand intérêt biologique, l'angiotensine II, ainsi que pour un polypeptide modèle qui simule sa chaîne principale (§ 3).

1 Schéma de structure et conformation : Dans de nombreuses études de taxinomie animale (cf e.g. dans ces Cahiers 1978, Vol III n° 1 les travaux de L. Bellier, & B. Abi Boutros; de V. Eisenmann et J.C. Turlot) le matériel de base est un ensemble de crânes : sur chacun de ces crânes, le spécialiste reconnaît les mêmes points de repère (pointe osseuse, extrémité d'une suture, limite d'une rangée dentaire...) entre lesquels on mesure des distances. De même, il est classique en chimie, de schématiser une molécule par un ensemble d'atomes ponctuels disposés suivant un graphe dont les arêtes sont les liens de valence : sous un unique schéma, sont comprises une infinité de réalisations spatiales possibles (différant quant à certaines distances entre atomes), ou conformations (cf § 2). La mécanique quantique permet en principe de reconnaître les conformations effectivement réalisées dans des conditions données et d'en calculer les probabilités relatives. Mais d'une part ces calculs sont très complexes; d'autre part la diversité des conformations envisagées est si grande, que pour en acquérir une vue d'ensemble il semble utile de recourir à l'analyse statistique multidimensionnelle.

Ici comme là (molécules ou crânes) c'est un même problème géométrique déjà traité dans ces cahiers (cf Ph. Bourgeois : Vol III n° 4; et addendum, Cahier Vol IV n° 2) le schéma de structure est donné par un ensemble fini I (ensemble de points de repère; ou ensemble des atomes...); une réalisation est un ensemble $M(I) = \{M^i | i \in I\}$, indicé par $i \in I$, de points M^i de l'espace euclidien de dimension S ; (si $\text{Card } I = n, M(I)$ est défini par un système M^I_J de $3n$ coordonnées: $M^I_J = \{M^i_j | i \in I; j = 1, 2, 3\} \in R^{3n}$; dans la suite on écrira indifféremment $M(I)$ ou M^I_J pour désigner une réalisation; et M^i ou M^i_J pour un point i) deux réalisations $M(I)$ et $N(I)$

(1) Dr G. Flamenbaum, C.E.A., CEN/Cadarache. Saint Paul-Lez-Durance.

(2) Mme E. Abillon, C.E.A., CEN/Saclay, Gif-sur-Yvette.

(3) J.P. Benzécri, Professeur de statistique. Université P. et M. Curie.

Les auteurs remercient Monsieur B. Fulman, Professeur de biologie physico-chimique à l'univ. P. & M. Curie et Monsieur J.L. de Coen, univ. Libre de Bruxelles Rhodes Saint GENESE. Belgique.

définissent une même *conformation* si l'une se déduit de l'autre par un déplacement global $g \in G$ (G , groupe des déplacements), i.e. si :

$$N(I) = gM(I) ; \forall i \in I : N^i = gM^i.$$

Bien que le *codage* de ces conformations, (c'est-à-dire le choix des systèmes de nombres par lesquels on les décrit), dépende essentiellement des contraintes physiques auxquelles doivent satisfaire les *réalisations* $M(I)$ (e.g. distance imposée entre M^i et $M^{i'}$; distance supérieure à un seuil entre M^i et $M^{i''}$; angle fixé à 10° près entre $M^i M^{i'}$ et $M^i M^{i''}$...), on donnera les principes généraux du codage avant de considérer les contraintes propres à l'analyse conformationnelle des molécules. Nous distinguerons deux voies : 1° Les coordonnées intrinsèques; 2° La distance interorbitale.

1° Les coordonnées intrinsèques : soit φ une application de R^{3n} dans R^m : cette application associe à toute *réalisation* $M(I)$ de I définie par le système M_J^I des $3n$ coordonnées des points M^i ($M_J^I = \{M_j^i \mid i \in I; j = 1, 2, 3\}$, cf *supra*) un système de nombres φ^h : $\varphi(M_J^I) = \{\varphi^h(M_J^I) \mid h = 1, \dots, m\}$. Ces nombres sont appelés des *coordonnées intrinsèques* si leur valeur est la même pour deux réalisations M et N définissant une même *conformation*; c'est-à-dire si φ ne change pas quand on soumet M à un déplacement global $g \in G$:

$$\forall M_J^I \in R^{3n}, \forall g \in G : \varphi(M_J^I) = \varphi(gM_J^I),$$

(où on a noté comme ci-dessus : $gM = gM_J^I = \{gM_j^i \mid i \in I\}$ la réalisation obtenue en faisant agir g sur M_J^I). Un système φ de coordonnées est dit *univoque* pour les conformations (on dira en bref: univoque) si deux réalisations M et N ayant même système de coordonnées $\varphi(M) = \varphi(N)$ définissent toujours une même *conformation*, i.e. si :

$$\forall M_J^I, N_J^I \in R^{3n} : \varphi(M_J^I) = \varphi(N_J^I) \Rightarrow \exists g \in G : N_J^I = gM_J^I.$$

Du point de vue de l'analyse des données, un système φ de m coordonnées intrinsèques permet de substituer à tout ensemble F de *conformations* de I , chacune définie par une *réalisation* particulière, un nuage de points de R^m : nuage que l'on soumettra aux algorithmes de classification et d'analyse factorielle.

2° La distance interorbitale : l'analyse des données s'applique non seulement aux nuages de points de R^m , mais aussi à tout ensemble fini F de points munis de masse et distance (on dit, en bref, un *triple* pour : {ensemble; masses; distances}) même si les points f de F ne sont pas représentés par des points $\varphi(f)$ d'un espace R^m : un tel *triple* est en effet la donnée requise pour une classification ascendante hiérarchique; on peut en faire l'analyse factorielle; et les algorithmes d'agrégation autour des centres variables (cf ce Cahier, [Boules Optimisées]) s'appliquent aussi, pourvu que les centres soient des points de F (on substitue donc à la recherche du centre de gravité d'une classe, celle du point de cette classe relativement auquel l'écart moyen de la classe est le plus faible). Donc si d'une manière ou d'une autre, on a pu définir et calculer une distance entre conformations (chacune de celles-ci étant données par n'importe laquelle de ses réalisations), il est possible de soumettre à l'analyse des données un ensemble de conformations.

Ceci dit nous rappelons (cf Lutz et Maïti in *Cahiers* Vol III n° 4) comment on définit une distance entre conformations, à partir des *orbites*

de leurs réalisations. Soit d'abord Dis une distance sur R^{3n} invariante par le groupe G des déplacements, i.e. telle que :

$$\forall M_J^I, N_J^I \in R^{3n}, \forall g \in G : \text{Dis}(M_J^I, N_J^I) = \text{Dis}(gM_J^I, gN_J^I).$$

(Par exemple, Dis est une somme, ou une somme pondérée de distances usuelles entre points homologues :

$$\text{Dis}(M, N) = \sum \{ \|M_J^i - N_J^i\| \mid i \in I \} ; \text{ ou mieux: } \text{Dis}^2(M, N) = \sum \{ \|M_J^i - N_J^i\|^2 \mid i \in I \} ;$$

cf Ph. Bourgeois, *Cahiers* Vol III n° 4 ; mais bien d'autres définitions sont possibles, sinon pratiques). La distance Dor entre deux conformations est définie comme le minimum de la distance Dis entre deux réalisations, M et N, définissant ces conformations ; on note :

$$\text{Dor}(M, N) = \inf \{ \text{Dis}(M, gN) \mid g \in G \} = \inf \{ \text{Dis}(gM, g'N) \mid g, g' \in G \} ;$$

le sigle Dor, *distance orbitale*, a été choisi parce que l'on appelle communément *orbite* l'ensemble des transformés d'un même M, par $g \in G$: ainsi les réalisations de I définissant la même conformation qu'une réalisation M constituent l'orbite de M : $\text{Orb}(M) = \{gM \mid g \in G\}$.

Il importe de noter ici que l'on peut prendre pour groupe G soit le groupe des déplacements (i.e. isométries directes) soit le groupe de toutes les isométries : ici il convient de se restreindre aux déplacements, si on ne veut pas identifier une structure avec son image dans un miroir (c'est ce qui est fait dans l'*addendum* au mémoire de Ph. Bourgeois).

Ph. Bourgeois (*loc. cit.*) a trouvé une méthode algébrique pour calculer exactement la distance orbitale entre deux réalisations d'un ensemble I non seulement dans R^3 , mais dans un espace euclidien de dimension quelconque : cette méthode n'était pas connue, lors de nos recherches sur la conformation de l'angiotensine II. A défaut de l'optimum que constitue la distance orbitale, on peut définir une distance entre conformations par toute formule Din_v de distance entre réalisations invariante non seulement par G mais par $G^2 = G \times G$, au sens suivant :

$$\forall M, N \in R^{3n}, \forall g, g' \in G : \text{Dinv}(gM, g'N) = \text{Dinv}(M, N) ;$$

(on notera la différence avec la formule d'invariance de Dis par G : ici interviennent deux déplacements distincts g et g'). Pratiquement on aboutit à des formules de distances Din_v, par des calculs fournissant une approximation de Dor (cf e.g. *in* Lutz et Maïti : amener M et N à avoir même 1° axe d'inertie ; et tourner N autour de cet axe pour minimiser Dis).

2 Structure des molécules : On l'a dit au § 1, le codage des objets spatiaux doit tenir compte des contraintes physiques propres au domaine étudié. Nous rappellerons donc brièvement au § 2.1 la structure des molécules organiques, afin d'en fixer au § 2.2 une description mathématique ; qui est à la base de l'analyse conformationnelle (§ 2.3). Ce schéma géométrique n'est toutefois qu'une vue simplifiée d'un système régi par la mécanique quantique à laquelle les spécialistes ont dû recourir pour préparer les données de nos analyses (§ 2.4).

2.1 Formules brutes ; formules développées ; stéréochimie : On part de trois corps le méthane ; l'éthanol (ou alcool éthylique usuel) ; et le chloro-bromo-iodoforme, composé dont la formule est choisie pour offrir un exemple très simple de molécule asymétrique. Puis on considère les polypeptides objet de la présente étude.

2.1.1 Le méthane : Formule brute CH_4 : la molécule comprend un atome de carbone et quatre d'hydrogène ; formule développée (cf figure 1), le carbone est relié à chaque hydrogène par un lien de valence long d'environ 1,1 Angström ; formule spatiale : le carbone est au centre d'un tétraèdre régulier dont les quatre atomes d'hydrogène occupent les sommets.

2.1.2 Le chlorobromoiodoforme : Formule brute CHClBrI ; formule développée : la molécule comprend un atome de carbone relié à quatre atomes différents ; formule spatiale : le carbone est relié aux quatre autres atomes par des liens d'inégales longueurs orientés suivant les mêmes directions que pour le méthane : on notera que les trois vecteurs $\overrightarrow{\text{CCl}}$, $\overrightarrow{\text{CBr}}$, $\overrightarrow{\text{CI}}$ forment sur la figure un trièdre direct : mais l'image dans un miroir de cette configuration est également possible ; on a alors un trièdre d'orientation inverse : c'est là un exemple aussi simple que possible de deux isomères optiques, images l'un de l'autre par rapport à un plan.

2.1.3 L'éthanol : Formule brute $\text{C}_2\text{H}_6\text{O}$; formule semi-développée $\text{CH}_3\text{-CH}_2\text{OH}$ (distinguant les deux atomes de carbone) ; formule spatiale : de chacun des deux atomes de carbone sont issus quatre liens de valence, orientés comme pour CH_4 ; l'angle des deux liens issus de l'atome d'oxygène OC, OH, est également fixe (120°) ; mais l'angle du plan φ_1 formé par les plans 123 et 234 (plans CCO et COH : pour le numérotage des atomes de 1 à 9, voir figure) n'est pas fixé ; et de même pour l'angle φ_2 des plans 123 et 127. A la différence de ce qui était le cas pour CH_4 et CHClBrI , la molécule est susceptible de se déformer tout en respectant le longueur des liens de valence ainsi que les angles entre liens issus d'un même atome ; elle peut prendre ainsi une infinité de formes dépendant de deux paramètres angulaires φ_1 , φ_2 ; ces angles sont appelés angles de torsion. On peut dire que le couple (φ_1, φ_2) constitue un système de *coordonnées intrinsèques* (au sens du § 1) ; Toutefois une ambiguïté se présente : rien ne distingue les rôles des atomes d'hydrogène numérotés 7, 8, 9 : l'angle φ_2 n'est donc défini qu'à 120° près. De telles indéterminations se rencontrent chaque fois que la structure est invariante par certaines permutations des atomes (cf *infra* § 2.2).

Toutefois une particularité importante se rencontre ici : du fait d'interactions complémentaires entre les atomes, non seulement les longueurs des liens de valence et les angles entre liens issus d'un même atome ont des valeurs fixées, mais encore un angle de torsion peut l'être également. C'est le cas pour l'angle φ_2 de l'éthanol qui (à 120° près) vaut soit 0° soit 60° : dans le premier cas on parle de conformation *éclipsée*, dans le 2^e cas de conformation *étoilée* : le sens de ces termes apparaît clairement sur la fig. 1' et sa légende. Dans les conditions usuelles, ces deux conformations coexistent ; et étant susceptibles de se transformer l'une en l'autre, elles sont en équilibre thermodynamique.

2.1.4 Protéines et polypeptides : Tout polypeptide est un polymère résultant de l'enchaînement d'acides aminés de formule générale :

$$\text{H}_2\text{N}-\underset{\text{R}}{\text{CH}}-\text{COOH}$$

méthane : CH_4	chlorobromochloroforme : CHCl_2BrI	éthanol : $\text{C}_2\text{H}_5\text{O}$

Figure 1 : Formules planes ou spatiales de trois composés organiques ; les longueurs des liens de valence sont indiquées en Angström ($10^{-4} \mu\text{m}$) ; sur la formule de l'éthanol on a numéroté les atomes de 1 à 9.

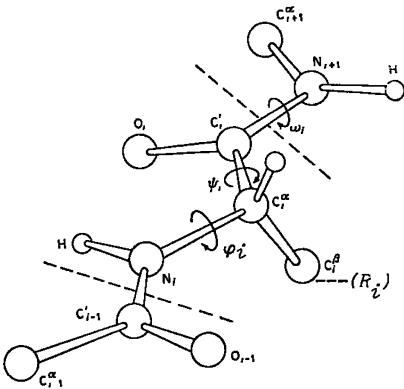


Figure 2 : formule spatiale d'une protéine : on note C_i^{α} l'atome de carbone du groupe $\text{C}=\text{O}$ du i -ème amino-acide ; C^{α} le carbone immédiatement relié à celui-ci d'où part la chaîne latérale R_i ; C_i^{β} le premier atome de cette chaîne ; sur la figure le trièdre $(C_i^{\alpha}, C_i^{\beta}, C_i^{\alpha} C_i^{\beta}, C_i^{\alpha} N_i)$ est direct : on dit que l'acide est dans la configuration L ; c'est le cas quasi-général dans les protéines des êtres vivants (figure d'après Edsall et coll. ; 1966 ; J. of Mol. Biol. 15 ; pp 399-407).

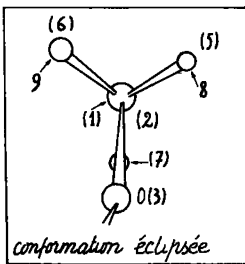
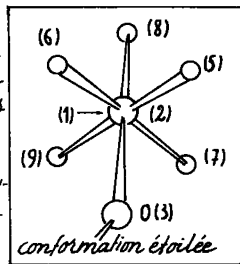
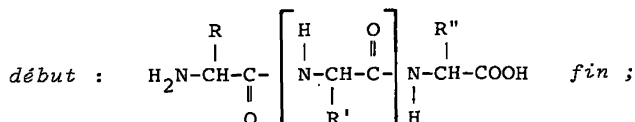


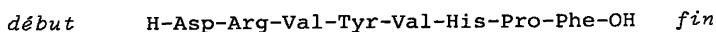
Fig.1: deux conformations de l'éthanol: la molécule est vue à partir d'un point situé dans le prolongement de l'axe (1) (2). On a marqué d'une flèche les atomes masqués ("éclipsés") sur une telle vue; l'aspect du dessin suffit à expliquer les noms attribués aux conformations: à droite, on a une étoile; à gauche, les 3 hydrogènes reliés au carbone 1 sont masqués, "éclipsés".



on a divers acides aminés dont 20 sont à la base de toutes les structures biologiques (cf *Cahiers* Vol IV n°2, [Code Gén] § 1). L'enchaînement de ces molécules se fait par liaison peptidique, suivant la formule:



sur cette formule semi-développée, on a indiqué suivant la convention usuelle le *début* et la *fin* de la chaîne ; le motif entre crochets (au milieu de la formule) représente une suite d'acides aminés généralement différents. Si la chaîne comprend e.g. une dizaine d'acides aminés on parle de polypeptide ; tandis que les protéines naturelles sont composées de plusieurs dizaines d'acides aminés. On écrit d'ordinaire la formule d'un polypeptide (ou d'une protéine) comme une suite de sigles à trois lettres (Gly pour glycine, Ala pour alanine etc.) représentant chacun un amino-acide (cf [Code Gén.] Vol IV n°2), ainsi la formule de l'Angiotensine II (bovine) est :



Sur la formule spatiale de la figure 2, on a figuré les trois angles de torsion φ_i , ψ_i , ω_i dont dépend (au niveau de son i -ème maillon) la conformation de la chaîne principale. D'autres angles (notés χ_{i1} , χ_{i2} ... χ_{ij} ...) sont nécessaires pour décrire la disposition de la chaîne latérale R_i du i -ème amino-acide : on ne les a pas indiqués ici.

Ajoutons trois remarques : (1°) En toute précision, il faut définir les angles de torsion comme des angles orientés de plans, (ainsi qu'on l'a fait au § 2.1.3 pour les angles de l'éthanol). Pour les conventions adoptées, nous renvoyons aux mémoires originaux.

(2°) L'angle de torsion ω est en fait à peu près fixé à l'une des valeurs 0° ou 180° parce que les 5 atomes $\{C_i^\alpha ; C_i ; O_i ; N_{i+1} ; C_{i+1}^\alpha\}$ définissent sensiblement un plan (dit plan de la liaison peptidique ; cf *infra* § 2.2.4) (3°) La notation C_i^β peut laisser croire que l'on rencontre toujours en cette position un atome de carbone : ce n'est pas le cas si le i -ème amino-acide est la glycine : on a alors là un hydrogène.

2.2 Description mathématique de la structure spatiale : Les exemples donnés au § 2.1, montrent que le schéma de structure d'une molécule ne se réduit pas à un ensemble fini I (ici, l'ensemble étiqueté des atomes, mais que l'on doit spécifier les longueurs des segments joignant certains atomes (liaisons chimiques), ainsi que les angles entre segments consécutifs. Dans ce cadre, un système de coordonnées intrinsèques se présente naturellement : ce sont les angles de torsion. Précisons les éléments de cette description, d'après l'exemple très simple de l'éthanol.

2.2.1 L'ensemble étiqueté des atomes

I : ensemble fini, muni d'une application e (ou *étiquetage*) prenant ses valeurs dans l'ensemble des corps simples (H, C, O, N, S, ... : hydrogène, carbone...). Dans le cas de l'éthanol ;

$$I = \{1, \dots, 9\} ; e(1) = e(2) = C ; e(3) = O ; e(4) = e(5) = \dots = e(9) = H$$

2.2.2 Le graphe des liaisons chimiques

L : ensemble fini de paires d'éléments de I, les arêtes du graphe, ou liens de valence ; chaque lien $l=(i,i')$ a une longueur déterminée $d(l) = d(i,i')$. Pour l'éthanol on a :

$$L = \{(1,2) (1,7) (1,8) (1,9) (2,3) (2,5) (2,6) (3,4)\} ;$$

(ici les extrémités de chaque lien sont données dans l'ordre du numérotage des atomes : e.g. (1,2) et non (2,1)) ; les longueurs des liens sont marquées sur la formule développée de la figure 1.

2.2.3 Les angles de valence : Ce sont les angles formés par deux liens issus d'un même atome : on note $\theta(i;i',i'')$ l'angle formé par les liens (i,i') et (i,i'') issus de l'atome i . Ces angles ne sont pas tous indépendants entre eux : e.g. si 4 liens sont issus d'un même atome i , cela fait 6 angles de sommet i ; mais il suffit d'en connaître 5 pour calculer le 6°. Pour l'éthanol les angles de courbure valent tous approximativement $109^\circ 28'$, (angle sous lequel est vue une arête à partir du centre d'un tétraèdre régulier) sauf $\theta(3;2,4)$ (angle de sommet de l'oxygène) qui vaut 120° .

$$\theta(1;2,7)\#\theta(1;2,8)\#\theta(1;2,9)\#\theta(1;7,8)\#\theta(1;8,9)\#\theta(1;9,7)\#109^\circ 28' \\ \theta(2;1,3)\#\theta(2;1,5)\#\theta(2;1,6)\#\theta(2;3,5)\#\theta(2;5,6)\#\theta(2;6,3)\#109^\circ 28'$$

2.2.4 Les angles de torsion (ou angles dièdres) : Dans les conditions physicochimiques de stabilité des molécules, les longueurs des liaisons et les angles de valence sont à peu près constants : au contraire les angles de torsion, n'ont pas de valeur fixe : et leurs variations simultanées engendrent la diversité des conformations. Soit $l = (i,i') \in L$, et (i,i_1) , (i,i_2) deux autres arêtes issues chacune d'une extrémité de l : on note $\varphi((i,i'))$; i_1, i_2 l'angle de dièdre ayant pour arête (i,i') formé par les plans (i,i',i_1) et (i,i',i_2) . On notera que (moyennant la convention internationale d'orientation telle que $0 < \varphi < \pi$, si le trièdre $\overline{ii'}$, $\overline{ii'_1}$, $\overline{ii'_2}$ est direct ; et $-\pi < \varphi < 0$ si le trièdre est inverse) on peut définir le signe de l'angle de torsion φ , (tandis que les angles de valence sont toujours compris entre 0 et π et rarement inférieurs à 90°).

Les angles de torsion ne sont pas indépendants entre eux : il suffit d'en donner un seul par arête (non terminale) : e.g. pour l'éthanol, les angles φ_1, φ_2 .

$$\varphi_1 = \varphi((2,3); 1,4) ; \varphi_2 = \varphi((1,2); 7,3).$$

Parfois la structure chimique impose que certains atomes soient coplanaires : l'angle de torsion n'est plus libre ; sa valeur est fixée à 0 (ou à π). Dans d'autres cas un angle de torsion n'est susceptible que de très faibles variations : par exemple dans les peptides (cf fig. 2 et l'appendice). Le véritable problème de l'analyse conformationnelle, est de reconnaître au voisinage de quels optima locaux se distribuent les combinaisons de valeurs des angles φ effectivement réalisées (cf § 2.4).

2.3 Symétries ; configurations ; conformations : On a vu sur deux exemples (§ 2.1.2 et § 2.1.4, cf figure 2, légende : configuration L) que des réalisations d'une même molécule peuvent différer par l'orientation d'un trièdre formé de liens issus d'un même atome. Tandis qu'au § 2.1.3 (éthanol) il est apparu qu'un angle de torsion peut être déterminé à $(2\pi/3)$ près si plusieurs atomes jouent dans la structure des rôles équivalents. On donnera à ces remarques un cadre général, en étudiant les symétries d'une molécule par permutation (§2.3.1) ; d'où une définition précise des conformations (§2.3.2) et des configurations (§ 2.3.3).

2.3.1 Symétries d'une molécule par permutation : Soit $\sigma \in I!$ une permutation de l'ensemble I (i.e. application biunivoque de I sur I). On dit que σ respecte la structure d'une molécule si σ laisse invariant l'étiquetage e , les liens et distances L, d , et les angles de courbure θ qui caractérisent cette structure (cf § 2.2) : i.e. si :

$$\forall i \in I : e(\sigma i) = e(i) ;$$

$$\forall l = (i, i') \in L : (\sigma i, \sigma i') \in L \text{ et } d(\sigma i, \sigma i') = d(i, i') ;$$

$$(i, i'), (i, i'') \in L \Rightarrow \theta(i; i', i'') = \theta(\sigma i; \sigma i', \sigma i'') ;$$

(éventuellement, cf § 2.2.4 si certains angles de torsion sont fixés par la structure, on demandera que σ les laisse invariants).

Les permutations σ respectant la structure forment un groupe $S\text{-}I!$: le groupe des symétries par permutation. Dans le cas de l'éthanol ce groupe comprend 12 permutations : c'est le produit direct du groupe (à deux éléments) des permutations des atomes $\{5, 6\}$ et du groupe (à 6 éléments) de permutations des atomes $\{7, 8, 9\}$.

2.3.2 Réalisations et conformations : Une réalisation $M_J^I \in \mathbb{R}^{3n}$ de l'ensemble I ($\text{Card } I = n$) dans l'espace \mathbb{R}^3 (cf § 1) est une réalisation de la structure d'une molécule, si elle satisfait aux conditions d'angles et distances propres à cette structure (cf § 2.2). On dit que deux réalisations M_J^I et N_J^I définissent une même conformation moléculaire s'il existe un déplacement direct $g \in G^+$ et une symétrie par permutation $\sigma \in S$ (cf § 2.3.1) tels que :

$$N_J^I = g\sigma M_J^I ; \text{ i.e. } \forall i \in I : N_{\sigma i}^I = gM_i^I ;$$

Il est équivalent de dire que si $(i, i'), (i, i_1), (i', i_2)$ sont trois liaisons de L on a :

$$\varphi((i, i') ; i_1, i_2) \text{ mesuré sur } M_J^I = \varphi((\sigma i, \sigma i') ; \sigma i_1, \sigma i_2) \text{ mesuré sur } N_J^I .$$

Nota Bene : en toute rigueur, la distance orbitale entre les conformations données par deux réalisations M et N doit se définir en tenant compte non seulement du groupe de déplacement G^+ , mais aussi de S , d'où la formule :

$$\text{Dor}(M, N) = \inf\{\text{Dis}(M, g\sigma N) \mid g \in G^+ ; \sigma \in S\}$$

2.3.3 Conformations et configurations : On dit que deux conformations appartiennent à une même configuration de la molécule, s'il existe pour ces conformations deux réalisations respectives M_J^I et N_J^I que l'on puisse relier par une suite continue (arc de courbe dans \mathbb{R}^{3n}) de réalisations de la molécule.

Il revient au même de dire que si M_J^I et N_J^I sont des réalisations quelconques de ces deux conformations, il existe σ dans S telle que M_J^I et N_J^I soient reliés par un arc dans \mathbb{R}^{3n} : une condition nécessaire pour cela est que les trièdres homologues formés par trois liens concurrents soient de même sens.

Soit ω une isométrie inverse (e.g. une symétrie par rapport à un plan ou un centre), on dit que ωM_J^I définit une *conformation inverse* de celle définie par N_J^I ; et que ces conformations appartiennent à des *configurations* inverses l'une de l'autre. En général, une conformation n'est pas égale à son inverse. Pour l'éthanol seule est égale à son inverse la conformation présentant un des atomes d'hydrogène {7,8,9} dans le plan principal (1,2,3); l'éthanol n'a qu'une seule configuration, qui est donc égale à son inverse. Tandis que le chlorobromoiodoforme, (cf § 2.1.2) est une molécule très simple qui n'a que deux conformations; mais celles-ci définissent deux configurations, inverses l'une de l'autre. Les corps optiquement actifs, découverts par Pasteur, sont ceux dont la configuration n'est pas égale à son inverse.

2.4 Le continuum des conformations : Du point de vue des contraintes géométriques énoncées au § 2.2, les conformations permises forment un espace topologique : les composantes connexes de cet espace sont les *configurations*. Dans les conditions de stabilité, une molécule peut changer de conformation, non de configuration : mais il est possible de passer continûment d'une conformation à une autre si celles-ci appartiennent à une même configuration. Toutefois ces conceptions géométriques sont loin de décrire exactement la réalité physique ; certaines conformations bien que satisfaisant aux contraintes d'angle et de distance du § 2.2, sont totalement irréalisables, parce qu'elles impliquent un recouvrement entre les nuages électroniques de deux atomes (on parle alors d'*empêchement stérique*, i.e. spatial) ; ou seulement une forte répulsion entre ces nuages. Pour juger de la possibilité d'une conformation, il faut calculer, au moins approximativement son niveau d'énergie d'après la répartition spatiale des charges électroniques (fonctions orbitales de la mécanique quantique ; cf § 3.2.1). Il apparaît alors que sur l'ensemble continu des conformations appartenant à une configuration donnée, seules certaines plages séparées les unes des autres et plus ou moins étendues sont à des niveaux d'énergie assez bas pour être réalisables. Le véritable objet de l'analyse conformationnelle est de découvrir ces plages (chacune centrée sur un minimum relatif de l'énergie) ; de préciser les passages de la molécule de l'une à l'autre ; etc. Ces passages ont probablement un rôle essentiel dans l'activité biologique des molécules. Acquérir une vue d'ensemble même simplifiée de l'espace des conformations, avec ses plages, plus ou moins étendues, et les voies de passage entre elles (cols de potentiel) serait donc très utile : c'est le but des analyses de données que nous avons tentées.

Il faut noter ici un conflit entre la terminologie du géomètre et celle du chimiste. Pour le géomètre, il est commode de considérer que chaque point du continuum des formes permises est une conformation. Pour le chimiste la conformation est (plutôt qu'un point de ce continuum) une plage étroite façonnée par un optimum énergétique local ; plage où la forme instantanée de la molécule a une certaine probabilité de se trouver ; mais à partir de laquelle des transitions sont possibles vers d'autres plages. De ce point de vue, il y a deux conformations pour l'éthanol (cf § 2.1.3 ; fig. 1' = $\varphi_2 \# 0^\circ$; $\varphi_2 \# 60^\circ$) ; il y a deux conformations de la liaison peptidique ($\omega \# 180^\circ$; $\omega \# 0^\circ$; cf § 2.1.4 ; § 2.2.4)..

Ajoutons qu'entre conformation et configuration la distinction n'est pas absolument claire. Pour le chimiste, il y a différence de configuration entre deux formes (ou deux plages) si entre celles-ci le passage est impossible dans les conditions usuelles de l'équilibre thermodynamique ; sinon ce n'est qu'une différence de conformation. Mais pour le géomètre une configuration se définit par la constance de certaines distances ; de certains angles ; (ainsi que des sens des trièdres). Cependant si l'on adopte cette dernière définition, on aura des configurations différentes entre lesquelles le passage est possible, et qui sont donc en équilibre thermodynamique : c'est le cas de la mutarotation des sucres. D'autre part il est concevable que l'empêchement stérique rende tout passage impossible entre deux conformations qui pourtant géométriquement parlant relèvent d'une même configuration.

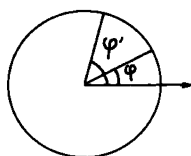
3 Analyse multidimensionnelle de l'ensemble des conformations d'une molécule

Dans toute analyse multidimensionnelle, on applique certaines méthodes à un tableau k_{IJ} , décrivant un ensemble I d'individus par un ensemble J de variables. Ici, choisir les variables, c'est décider du codage adopté pour décrire les conformations d'une molécule et calculer leurs distances mutuelles (§ 3.1). Un individu, s'identifie en principe à la description mathématique fixée pour lui : toutefois, il faut se restreindre à un échantillon de conformations physiquement réalisables, c'est-à-dire dont l'énergie est assez basse (§ 3.2).

Quant aux méthodes, renvoyant à un autre article l'exposé d'un algorithme de classification conçu pour les présentes données (cf ce Cahier [Boules Optimisées]) ; on se bornera à rendre compte avec quelques détails d'une application de l'a. des correspondances (§ 3.3).

3.1 Codage des conformations : Appliquant à la structure des molécules décrite au § 2, les principes géométriques énoncés au § 1, on décrit ci-dessous quatre systèmes de codage ; puis on compare ces systèmes

3.1.1 Les angles de torsion : La définition générale de ces angles est donnée au § 2.2.4 ; pour un polypeptide, on utilise les angles $\varphi, \psi, \omega, \chi$ expliqués au § 2.1.4 et sur la figure 2. Admettons que les angles ω sont bloqués (ils le sont en effet à 10° près, cf § 2.1.4 ; or il est impossible d'explorer l'espace conformationnel, en tenant compte de si petites variations) : il reste, dans le cas de l'angiotensine II, 30 angles $\{\varphi, \psi, \chi\}$, (dont le détail est dans la thèse). Ces angles forment (au sens du § 1, 1°) un système de coordonnées intrinsèques de dimension 30, que l'on notera aussi $\{\varphi_a | a = 1, \dots, 30\}$. Aux rotations près de quelques méthyles $-\text{CH}_3$, autour de leur liaison de valence (rotations dont on peut faire abstraction) c'est là un codage univoque (i.e. permettant de distinguer dans tous les cas deux conformations distinctes ; cf § 1). En vue d'une classification automatique ou d'une analyse de triple (cf § 1, 2°) il faut traduire en distances les différences de coordonnées.



Entre deux points M et M' du cercle de rayon 1 définis par une abscisse angulaire unique φ, φ' , la distance au carré est : $4 \sin^2((\varphi - \varphi')/2)$ (expression invariante par rotation ; et qu'il suffit de vérifier dans le cas où φ' est nul). Pour deux configurations, définies par une suite d'angles φ_a, φ'_a , on posera en général :

$$d^2 = \sum \{p_a \sin^2((\varphi_a - \varphi'_a)/2) | a = 1, \dots, 30\}.$$

Les coefficients p_a pouvant être choisis selon l'influence d'une variation de φ_a sur la forme générale de la molécule (e.g. une rotation au milieu de la chaîne principale, influe plus qu'une rotation à l'extrémité d'une chaîne latérale...). Il est également possible de calculer la distance comme une somme pondérée de valeurs absolues $|\sin(\varphi_a - \varphi'_a)/2|$.

3.1.2 Les angles de Flory : Considérons, d'après Lewis & coll (1971 : Proc. Acad. Sc. U.S.A. ; 68 ; 2293-2297) deux exemples de tournant β dans la chaîne d'un polypeptide : globalement, les deux cas se ressemblent,

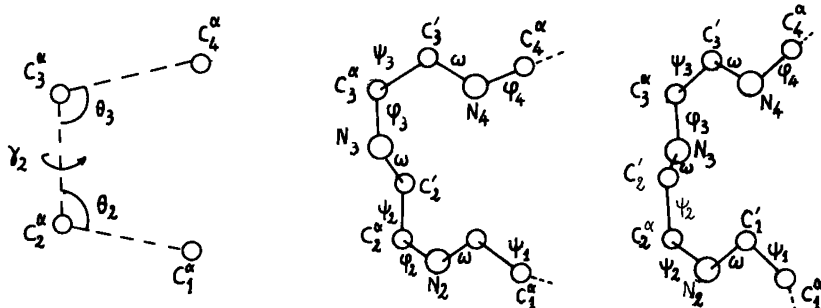


Figure 3 : à droite, deux modèles (I et II) de tournant β dans la chaîne d'un polypeptide, réalisés par le jeu d'angles de torsion différents ; à gauche schéma de la chaîne des carbones α .

mais ils sont réalisés par des angles de torsion tout à fait différents : pour I : $\{\varphi_2 ; \psi_2 ; \varphi_3 ; \psi_3\} = \{-50^\circ ; -30^\circ ; 100^\circ ; 30^\circ\}$; pour II on a $\{-30^\circ ; 90^\circ ; 120^\circ ; -30^\circ\}$. La similitude globale des formes ne requiert donc pas nécessairement une proximité entre les angles de torsion. On aura des coordonnées intrinsèques plus directement interprétables que les φ, ψ , en codant par courbure et torsion la chaîne principale réduite à ses atomes C_1^α (cf fig. 2).

3.1.3 Les distances interatomiques : Comme on le fait pour un crâne (cf § 1) on peut décrire une molécule par le système des distances entre un certain nombre de points de repère : ici ces points seront évidemment non des pointes osseuses ni des sutures... mais des atomes. La conformation d'une molécule d'Angiotensine II serait parfaitement déterminée par les positions relatives d'une trentaine d'atomes. Dans un but de simplification, on a utilisé des représentations comportant 20 ou 21 atomes, en veillant à préciser, outre la chaîne principale, l'orientation des résidus aromatiques des aminoacides 4, 6 et 8 (Tyr⁴, His⁶, Phe⁸). Pour n atomes, on a $n(n-1)/2$ distances interatomiques : ici $20(19)/2 = 190$. Compte-tenu des distances fixées par les liaisons chimiques il reste 180 paramètres. C'est là un nombre de variables acceptables dans une analyse de correspondance (cf 3.3).

Le lecteur attentif à la stéréochimie, objectera ici qu'en toute rigueur un système de distances ne définit une forme qu'à une symétrie près. Mais qu'en résulte-t-il pour nous ? Simplement ceci : à chaque point, ou à chaque plage de notre espace conformationnel (représenté e.g. dans l'espace R^4 des quatre premiers facteurs issus d'une analyse de correspondance) il correspond en réalité deux conformations ou deux plages symétriques l'une de l'autre ; les plages sont bien reconnues ; bien situées les unes relativement aux autres ; et les passages entre elles aussi : pourvu seulement que ces passages ne relient pas en quelques étapes une forme à sa symétrique (dans un miroir). Cette éventualité (qui est fort rare hormis les cas analogues à la mutarotation des sucres évoquée ci-dessus, § 2.4 *in fine*) ne serait pas décelée, du moins au premier examen, sur les résultats d'une analyse du seul tableau des distances interatomiques.

3.1.4 La distance interorbitale : Une molécule étant assimilée (cf § 3.1.3) à un système simplifié de 21 atomes, on calcule la distance entre deux conformations définies par des réalisations M_J^I et N_J^I de la molécule, en soumettant N_J^I à un déplacement g minimisant la somme $\sum \{d^2(M_J^i, g N_J^i) | i \in I\}$, (cf § 1, 2°). La méthode de Ph. Bourgeois permet d'effectuer exactement cette minimisation ; toutefois on s'est borné à des calculs approchés.

3.1.5 Critiques et comparaisons : Au § 2.3.1 on a défini le groupe S des symétries d'une molécule par permutation ; ces symétries interviennent e.g. dans le calcul de la distance interorbitale (§ 2.3.2) : or on n'en n'a pas tenu compte ici. C'est que ces permutations n'affectent que quelques atomes d'hydrogène, (voire des éléments de chaîne latérale) qui ont disparu de notre description simplifiée.

Les angles de torsion changent de signe par une isométrie inverse ω (cf § 2.3.3) : mais les distances sont inchangées. En conséquence, si l'on utilise comme coordonnées intrinsèques les distances interatomiques, deux conformations inverses l'une de l'autre sont considérées comme identiques : le véritable espace conformationnel est donc un revêtement d'ordre 2 de celui qu'on représente en ne codant que des distances. Toutefois notre but étant de classer et de déployer spatialement certains ensembles de conformations dont on a reconnu l'importance physique (cf 3.2) la représentation par les distances peut suffire.

On a introduit les angles de Lewis au § 3.1.2 parce que la similitude globale de deux conformations est compatible avec de grandes différences se compensant mutuellement ; de telles compensations sont également possibles dans le système des angles de Lewis. La distance calculée d'après les variables angulaires n'est donc pas une mesure fidèle de l'écart global entre les formes : de ce point de vue, l'usage des distances interatomiques ou de la distance interorbitale est beaucoup plus sûr (c'est par comparaison avec l'analyse factorielle du tableau de correspondance des distances interatomiques, cf § 3.3, que l'on a décelé les imperfections du système des angles de Lewis). Pratiquement (cf § 3.2) les variables angulaires servent à repérer exactement une conformation et à calculer par des formules trigonométriques les coordonnées des atomes : de là on passe aux distances interatomiques (ou à la distance interorbitale) avant de procéder à l'analyse des données.

En définitive compte-tenu du coût des calculs de distance interorbitale, on recourra principalement aux calculs de distance interatomique (cf § 3.3 conclusion).

3.2 Exploration de l'espace conformationnel : Comme on l'a dit au § 2.4, seules nous intéressent les conformations dont le niveau d'énergie est voisin du minimum : l'exploration proprement dite (§ 3.2.2) requiert donc un calcul approché de l'énergie (§ 3.2.1).

3.2.1 Calcul de l'énergie d'une conformation : Des calculs complets dans le cadre de la mécanique quantique (cf e.g. Pullman *in Adv. in Protein Chem.* 38 ; 1974) ne sont faisables que pour des molécules plus petites que l'Angiotensine II. Dans la pratique on part d'un modèle de sphères dures (modèle assignant à chaque atome un nuage électronique de rayon déterminé, tout empiètement entre nuages étant interdit : à ce modèle correspondent les assemblages de billes de bois, utilisés tant dans l'enseignement que dans la recherche, pour représenter les molécules) pour décider *a priori* de la possibilité d'une conformation ; puis on évalue l'énergie par des formules semi-empiriques comportant divers termes (électrostatique ; van der Waals ; torsion ; liaison hydrogène). Dans cette voie, ouverte par Ramachandran (1963), on citera les travaux de Liquori, Lewis, Ralston...

3.2.2 Exploration des conformations de basse énergie : Comme dans toute recherche d'optimum, on pense d'abord à balayer l'espace entier en faisant varier simultanément pas à pas toutes les coordonnées : une telle exploration est impossible ici, car avec 30 (cf § 3.1.1) ou seulement 20 variables angulaires prenant chacune seulement $2^3 = 8$ valeurs différentes (ce qui est très grossier : $\Delta\varphi = 45^\circ$) on a $2^{60} \approx 10^{18}$ combinaisons pour chacune desquelles il faudrait calculer une énergie...

Voici qu'elle est la voie suivie : d'abord on choisit dans l'espace conformationnel quelques points dits *conformations d'attaque* ; puis à partir de chacun de ces points on procède à la recherche d'un optimum local de l'énergie par un programme itératif usuel (gradient ; simplexe..). Du choix des conformations d'attaque, dépend essentiellement l'efficacité de l'exploration. Les modèles classiques de structure des protéines (en hélice, en feuillet) sont un terme de référence essentiel. Pour plus de diversité, on a recours à la méthode du *stéréoalphabet* de Liguori (*in Quart. Rev. Biophys.* 2, 65-79 ; 1969) : en bref pour engendrer les conformations favorables d'un polypeptide, on enchaîne des conformations de chaque acide aminé correspondant à un optimum énergétique pour les unités dipeptidiques.

3.3 Exemples d'analyses : A deux jeux de données, (§ 3.3.1), on a appliqué un ensemble de méthodes (§ 3.3.2). Nous ne rendons compte en détail ici que d'une seule analyse de correspondance (§ 3.3.3) qui suggérera la conclusion de ce travail.

3.3.1 Les données analysées : Elles concernent deux molécules, l'Angiotensine II et l'octapeptide modèle (Ala)₆-Pro-Ala. Le choix des conformations étant fait comme on l'a dit au § 3.2.

L'Angiotensine II : on a 11 conformations d'attaque désignées par les lettres de A à K : les 9 premières sont celles de plus basse énergie fournies par le stéréoalphabet. Les deux dernières, J et K sont des conformations *étendues* l'une ayant les angles χ_1 tous égaux à 0° ; l'autre à 180° . A partir de chacune de ces 11 conformations, on en a obtenu 6 à 8 autres par itération (Abillon, 1976 ; Tournarie 1973) . Les conformations de chaque famille sont numérotées A_1 , A_2 , A_3 ... etc. , à partir de la conformation d'attaque A_1 .

Le polypeptide (Ala)₆-Pro-Ala : qui peut être considéré comme un modèle du squelette de l'angiotensine, a fait l'objet d'une étude conformationnelle approfondie dans le laboratoire de J.L. De Coen (cf De Coen & Ralston). Les auteurs ont engendré les conformations données par le stéréoalphabet, et retenu 1442 conformations dont l'énergie dépassait le minimum absolu de moins de 8 kcal (par mole).

3.3.2 Les méthodes appliquées : Au § 3.1 on a proposé quatre formules de distances (dont le calcul peut être éventuellement modifié par l'introduction de valeurs absolues : cf § 3.1.1 *in fine*): toutes ces formules se prêtent à la classification ascendante hiérarchique, ainsi qu'à la partition en boules optimisées (avec divers choix du rayon). Le tableau des distances interatomique (cf § 3.1.3) se prête directement à l'analyse des correspondances : et l'on peut simultanément utiliser la distance du χ^2 pour une C.A.H., ou une partition en b. o.. On a donc multiplié les essais ce qui nous a permis d'acquérir une vue d'ensemble sûre des données dont nous disposions.

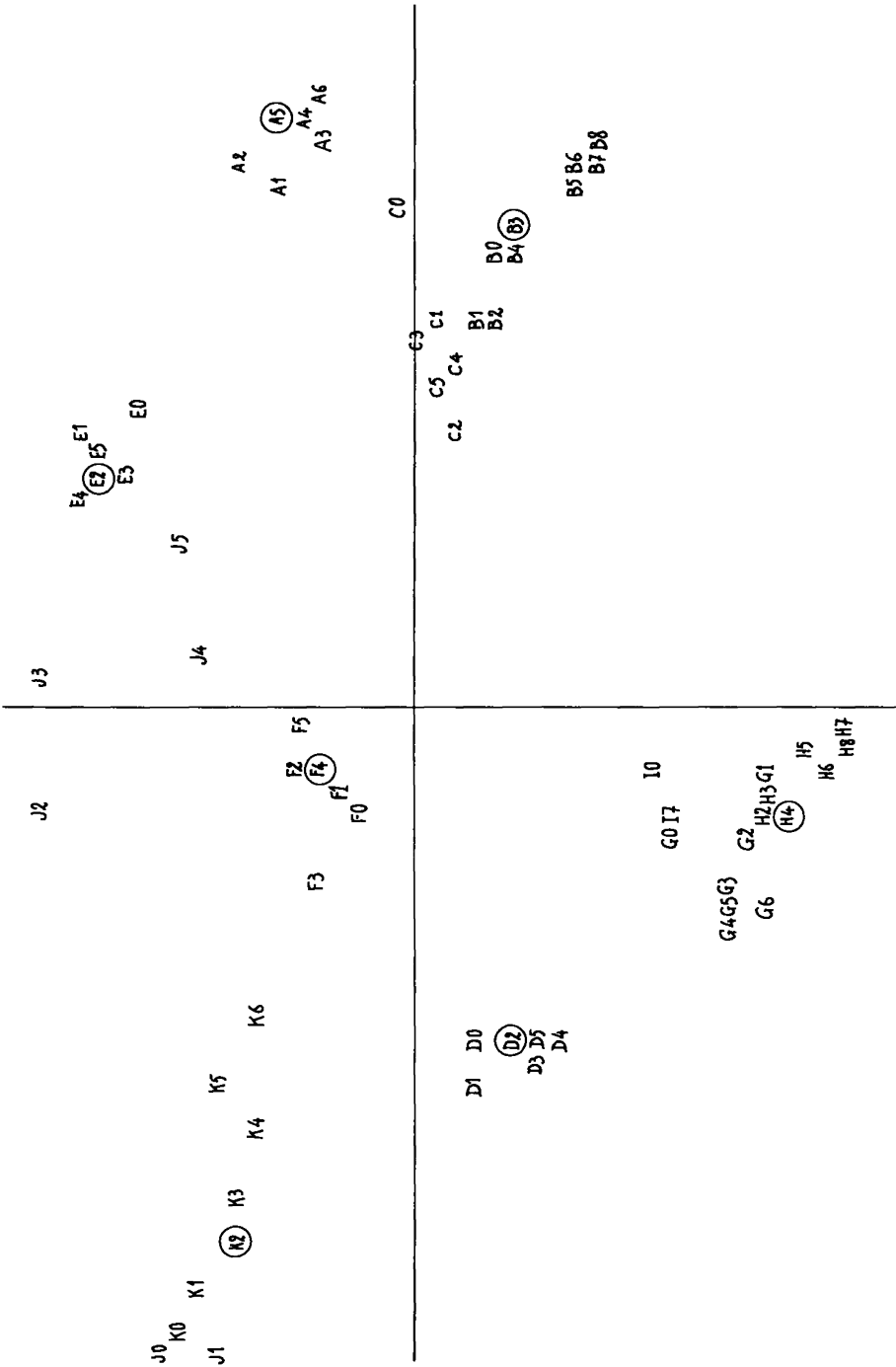


Figure 4. projection dans le plan principal d'inertie de 73 conformations de l'angiotensine II

3.3.3 Analyse de correspondance du tableau des distances interatomiques pour l'Angiotensine II : Tableau k_{IJ} : 72 conformations x 180 distances ; dans une telle analyse la représentation simultanée des deux ensembles I et J (conformations ; et distances interatomiques servant à décrire celles-ci) aide grandement à l'interprétation : chaque groupe de conformations va avec des caractères distances j qui prennent pour celles-ci des valeurs élevées ; et s'oppose aux caractères représentant des distances faibles. Dans la présente interprétation, on considère trois facteurs totalisant les 3/4 de l'inertie du nuage : $\tau_1 = 38,5\%$,

$\tau_2 = 21,5\%$; $\tau_3 = 14,5\%$). Afin de ne pas alourdir le texte par une explication détaillée des notations suivies (une lettre ou un chiffre par atome retenu ; donc un diagramme par distance e.g. 2U, ou Q6) on a seulement figuré l'ensemble I et on fait référence aux aminoacides sans détailler les positions des atomes pris un à un.

Sur le premier axe, on a l'opposition entre les familles des conformations étendues J, K (fortes valeurs des distances 1W, OW, 1V, PV..) aux familles présentant un tournant U (A,B,C) qui leur confère une conformation plus compacte. On peut dire également que l'axe principal d'inertie oppose les conformations ayant une direction principale d'inertie nettement marquée aux conformations d'un type nettement plus sphérique. Rappelons que ce phénomène ne pouvait pas être attendu avec certitude *a priori*, puisque la représentation des conformations n'était pas fondée sur les atomes de la chaîne principale, mais sur ceux des chaînes latérales.

Le second axe oppose notamment les familles H et E. Pour ces familles, on lit sur le premier axe qu'il s'agit de conformations dont la chaîne principale forme un U, les deux valines Val³ et Val⁵ étant situées à chacune de ses deux articulations. Le second axe précise la position des chaînes latérales autour de la chaîne principale, notamment la position des résidus Arg², Tyr⁴, His⁶, Phe⁸. En effet les corrélations et oppositions indiquées en particulier sur le nuage J des caractères, nous enseignent que deux types de tournant U sont à distinguer : d'une part ceux pour lesquels Tyr et Phe sont proches ; d'autre part ceux pour lesquels la distance Arg-His est réduite. Ce qui suggère donc qu'en présence d'un tournant U on a l'alternative suivante : ou l'espace intérieur compris entre les deux branches du U est occupé par Tyr et Phe, simultanément, ou alors par Arg et His. D'autres indications confirment cette assertion, comme celle-ci par exemple (également lisible d'après l'ensemble J), l'histidine rejetée de l'espace intérieur du tournant, se rapproche de Val⁵.

La famille F s'oppose à toutes les autres familles sur le troisième facteur, et est la seule à n'être pas correctement représentée dans le plan des deux premiers axes. Une excellente représentation de l'espace conformationnel engendré par les 73 conformations étudiées ici serait donc constituée par la disposition de la figure 4, les points de la famille F étant élevés verticalement à partir de leurs projections d'une cote égale à la demi-largeur de la figure. L'isolement de F est dû au fait que le tournant U esquissé par les résidus Val³ à Phe⁸ est escamoté de Asn¹ à Val³, tandis que Arg² tend à pallier cet effet. Cela étant, les positions des chaînes latérales échappent à l'alternative précédemment décrite, puisque toutes les chaînes latérales se trouvent être à l'extérieur du tournant.

Outre ces considérations topologiques, l'étude précédente apporte les informations suivantes :

On sait que les conformations sont issues d'un calcul d'optimisation. La figure 4 montre la *convergence éventuelle du processus itératif* dans lequel on est engagé. Effectivement, s'il y a convergence, les dernières conformations d'une même famille doivent être confondues (comme pour la famille A, ou A6, A7, A8, sont identiques), ou tout au moins proches l'une de l'autre (comme cela se produit pour B5, B6, B8, pour H5, H6, H7, H8, et pour les familles D, E, F, dont les projections de divers éléments sur les quatre premiers axes factoriels forment un ensemble compact). Par contre, il est sûrement souhaitable de poursuivre l'effort de minimisation de l'énergie pour les familles G, J et K, dont les éléments sont très dispersés en projection sur le plan principal.

Le plan de la figure résume donc deux critères principaux pour la description de la forme d'un polypeptide linéaire : forme de la chaîne principale pour l'axe 1, disposition des résidus autour de la chaîne latérale pour l'axe 2. Mais la distribution dans ce plan des différentes conformations, leur évolution au cours d'itérations successives montre que cet espace n'est pas figé : ainsi il apparaît des zones de non stabilité conformationnelle (zones occupées par J0, J1, J2, J3 ou par la famille K), des zones interdites, ou des zones où se regroupent des éléments d'une même famille, véritables flots qui correspondent aux puits énergétiques.

3.4 Conclusion : Devant la nouveauté du domaine étudié, on a dans cette étude essayé de nombreux codages et conjugué plusieurs méthodes d'analyse : il est très satisfaisant que malgré les contraintes physiques particulières propres à ces objets de l'espace, les conformations moléculaires, aussi bien que les crânes des vertébrés se prêtent à une description quasi-universelle : prendre pour variables j les distances entre un certain nombre de points de repères (ici des atomes).

Mais avant d'organiser par l'analyse des données l'espace sous-tendu par l'ensemble des conformations que peut adopter une molécule, il faut explorer cet ensemble : engendrer judicieusement de nombreuses conformations, et calculer leur énergie avec une précision suffisante : l'analyse statistique multidimensionnelle n'intervient qu'après que de telles données ont été préparées.

A ce prix, il sera sans doute possible d'acquérir par l'étude de nombreuses molécules des vues générales nouvelles sur la dynamique des conformations : l'analyse factorielle donnera les cartes de base ; et la classification automatique servira à cerner les zones principales ; à découvrir des détails indiscernables autrement ; à distinguer certains recouvrements inévitables sur une figure plane.

La présente étude étant limitée à deux molécules parentes n'a pas apporté de résultats bruts inattendus ; mais nous avons le sentiment, en arrêtant ce travail, que tout n'a pas été dit ; et que s'ouvre devant l'analyse conformationnelle un champ nouveau d'observation et de réflexion.

Note bibliographique (J.-J. Bourgarit) : En appendice à un article consacré à la comparaison de différentes structures de protéines homologues (les chymotrypsiques) (cf J. Mol. Biol. Vol. 128 pp 49-76) A.D. Mc Lachlan traite de l'ajustement de deux nuages de points sous le titre "*Least-squares fitting of two structures*". Il s'agit en effet de définir un opérateur de rotation R rendant minima la somme :

$$\sum_n w_n (R b_n^+ - a_n^+)^2,$$

où les w_n sont des poids fixés ; et les a_n^+ , b_n^+ les positions d'atomes homologues appartenant respectivement à l'ensemble A (la protéine A) et à l'ensemble B (la protéine B) ; les deux ensembles étant rapportés chacun à son centre de gravité.

APPENDICE

La chaîne polypeptidique - Notations -

Si on examine une chaîne polypeptidique, on remarque que sa géométrie peut être modifiée si l'on fait varier :

- les longueurs des liaisons covalentes
- les valeurs des angles de valence
- les angles de torsion ω , φ et ψ de la chaîne principale
- les angles de torsion χ des chaînes latérales.

L'angle diédral ω décrit la liaison peptidique. Il peut adopter en principe les deux valeurs de 180° (trans) et 0° (cis) bien qu'en général ce soit la première qui, pour des raisons énergétiques, est favorisée. Pour les liaisons peptidiques impliquant ce résidu proline ou hydroproline, la géométrie *cis* peut atteindre parfois des pourcentages élevés. Qu'elle soit *cis* ou *trans* la liaison peptidique peut être considérée comme plane, ce qui favorise la circulation des électrons délocalisés : dans les protéines, les variations de ω sont effectivement faibles ($\Delta\omega < 10^\circ$). De même, les barrières d'énergie qui contraignent les angles de valence et les liaisons covalentes à rester voisines de leurs positions d'équilibre sont d'un ordre de grandeur tel que ces déformations sont quasiment interdites.