

Nœuds électrostatiques, monobrins d'ADN et minimisation de fonctionnelles

François Dubois^{*}

Résumé

Nous proposons deux fonctionnelles discrètes dont les minima pourraient être reliés d'une part à la classification des nœuds dans l'espace tridimensionnel et d'autre part aux formes possibles des monobrins d'acide ribonucléique. Dans les deux cas, il est préférable de poser un problème pénalisé pour envisager une éventuelle résolution numérique qui sera de toutes façons extrêmement coûteuse.

Plan

- I- Théorie des nœuds
 - I.1) Introduction
 - I.2) Energie électrostatique d'un lacet discret
 - I.3) Nœuds
 - I.4) Approximation numérique
- II- Forme naturelle d'un monobrin d'ADN
 - II.1) Introduction
 - II.2) Potentiel d'interaction
 - II.3) Pénalisation
- III- Remerciements
- IV- Références bibliographiques

^{*} Conservatoire National des Arts et Métiers, Paris. Document "ASCI" du 14 septembre 1998, édition 25 février 2007.

-I- THÉORIE DES NŒUDS.

I.1) Introduction.

• On entend ici par lacet une application continue γ du cercle S^1 à valeurs dans \mathbb{R}^3 . Le problème posé par les mathématiciens (voir par exemple Adams [Ad94]) est de classifier les nœuds, c'est à dire de définir des classes d'équivalence à isotopie près dans l'espace L des lacets :

$$(1.1) \quad L = C(S^1, \mathbb{R}^3).$$

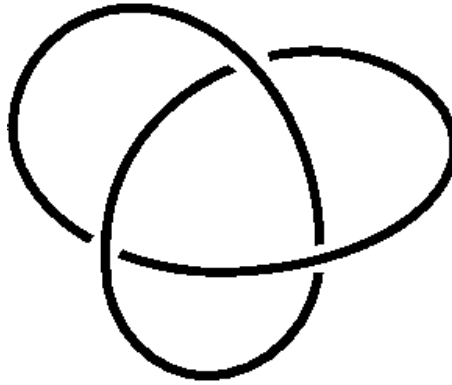


Figure 1 Nœud de trèfle.

• L'intuition courante nous dit qu'un point (l'application γ est une constante) n'est pas un cercle (l'application γ est par exemple l'injection $S^1 \rightarrow S^1 \times \{0\} \subset \mathbb{R}^3$), qui est lui même différent d'un nœud de trèfle (Figure 1). Nous nous proposons de tenter de caractériser une classe de nœuds par une "forme idéale", en suivant une démarche voisine de Katritch *et al* [Ka96]. Nous nous plaçons dans un cadre discret en excluant *a priori* le cas d'un lacet réduit à un point, laissant pudiquement ouvert le cas du passage à la limite continue.

• On désigne par $N \geq 3$ un entier et par X_N l'ensemble des suites $x = (x_0, x_1, \dots, x_N)$ de points de \mathbb{R}^3 deux à deux distincts sauf pour les extrémités où nous supposons $x_0 = x_N$:

$$(1.2) \quad \left\{ \begin{array}{l} x \in X_N = \left\{ (x_0, x_1, \dots, x_N) \in (\mathbb{R}^3)^{N+1}, x_0 = x_N, \forall (i, j) \in \mathbb{N}^2, \right. \\ \left. \left((0 \leq i < j \leq N) \text{ et } ((i, j) \neq (0, N)) \right) \Rightarrow (x_i \neq x_j) \right\}. \end{array} \right.$$

I.2) Energie électrostatique d'un lacet discret.

• Par définition, un N -lacet de longueur ℓ est la donnée de $y \in Y_N^\ell \subset X_N$ vérifiant une condition supplémentaire d'égalité des distances entre les sommets :

$$(1.3) \quad y \in Y_N^\ell = \left\{ x \in X_N, \forall j \in \{0, 1, \dots, N-1\}, |x_{j+1} - x_j| = \frac{\ell}{N} \right\}.$$

Pour chercher une forme qui soit la plus “tendue” possible, il est assez naturel d’imaginer que le lacet discret est formé de particules chargées de même signe qui se repoussent deux à deux tout en gardant la même distance entre deux charges consécutives et d’introduire l’énergie électrostatique $W_N^\ell(y)$ d’un lacet $y \in Y_N^\ell$ de longueur ℓ selon la relation algébrique classique :

$$(1.4) \quad Y_N^\ell \ni y \longmapsto W_N^\ell(y) = \frac{1}{4\pi} \frac{\ell}{N^2} \sum_{1 \leq i < j \leq N} \frac{1}{|x_i - x_j|} \in \mathbb{R}^+.$$

• Nous justifions la relation (1.4) par l’argument qui suit. Considérons les N points $x_j (j = 1, \dots, N)$ comme des charges ponctuelles de même signe

$$(1.5) \quad q = \frac{\sqrt{\ell}}{N}.$$

Nous plaçons la première charge q au point x_1 et nous cherchons le champ électrique E_1 créé par cette charge. Il s’agit d’un problème électrostatique qui s’écrit simplement

$$(1.6) \quad \begin{cases} \operatorname{div} E_1 = q \delta(x - x_1), & x \in \mathbb{R}^3 \\ \operatorname{rot} E_1 = 0, & x \in \mathbb{R}^3 \\ E_1 \longrightarrow 0, & |x| \rightarrow \infty. \end{cases}$$

La solution de ce problème est classique (voir par exemple Jackson [Ja62]) :

$$(1.7) \quad E_1 = -\nabla \left(\frac{q}{4\pi |x - x_1|} \right) \equiv -\nabla \varphi_1.$$

• L’énergie électrostatique V_1 pour amener une seconde charge depuis l’infini jusqu’à une distance

$$(1.8) \quad a = \frac{\ell}{N}$$

de la première à une position x_2 vaut $V_1 = -\int_\infty^{x_2} q E_1 \bullet ds$ et on a le calcul élémentaire suivant :

$$(1.9) \quad \begin{aligned} V_1 &= q \int_\infty^{x_2} d\varphi_1 = q (\varphi(x_2) - \varphi(\infty)) \\ V_1 &= \frac{q^2}{4\pi} \frac{1}{|x_2 - x_1|}. \end{aligned}$$

Ayant placé k charges ponctuelles aux points x_1, x_2, \dots, x_k , le champ électrique E_k est solution du problème de Poisson suivant

$$(1.10) \quad \begin{cases} \operatorname{div} E_k = q \sum_{j=1}^k \delta(x - x_j), & x \in \mathbb{R}^3 \\ \operatorname{rot} E_k = 0, & x \in \mathbb{R}^3 \\ E_k \longrightarrow 0, & |x| \rightarrow \infty \end{cases}$$

dont la solution

$$(1.11) \quad E_k = -\nabla \varphi_k$$

est une simple combinaison linéaire de k problèmes étudiés aux relations (1.6) et (1.7) :

$$(1.12) \quad \varphi_k(x) = \frac{q}{4\pi} \sum_{j=1}^k \frac{1}{|x - x_j|}, \quad x \in \mathbb{R}^3.$$

• L'énergie V_k nécessaire pour amener une $(k + 1)$ ième charge q à une distance donnée des k précédentes est, par un raisonnement analogue à celui conduisant à la relation (1.9), égal à l'expression suivante :

$$(1.13) \quad V_k(x) = q \varphi_k(x_{k+1}) = \frac{q^2}{4\pi} \sum_{j=1}^k \frac{1}{|x_{k+1} - x_j|}, \quad x \in \mathbb{R}^3.$$

L'énergie totale pour amener $(N - 1)$ charges auprès de la première dans la configuration y avec $y \in Y_N^\ell$ (y est un N -lacet de longueur ℓ) vaut donc

$$W_N^\ell(y) = \sum_{k=1}^{N-1} V_k = \frac{q^2}{4\pi} \sum_{k=1}^{N-1} \sum_{j=1}^k \frac{1}{|x_{k+1} - x_j|},$$

relation qui est identique à la relation (1.4) compte tenu de l'hypothèse (1.3) faite sur la charge électrique.

I.3) Nœuds.

• Nous posons le problème de chercher un point \bar{y} qui minimise localement l'énergie dans l'espace des N -lacets de longueur ℓ :

$$(1.14) \quad \begin{cases} \bar{y} \in Y_N^\ell \\ \exists \eta > 0, \forall y \in (\text{boule de centre } y \text{ et de rayon } \eta), W_N^\ell(\bar{y}) \leq W_N^\ell(y). \end{cases}$$

Notons que nous n'avons pas défini ici de topologie dans l'espace Y_N^ℓ mais qu'il faudra le faire. Nous appelons ensuite chemin admissible entre deux N -lacets discrets x et y de longueur ℓ ($(x, y) \in (Y_N^\ell)^2$) une application continue $\rho : [0, 1] \rightarrow Y_N^\ell$ telle que

$$(1.15) \quad \rho \in C([0, 1], Y_N^\ell), \rho(0) = x, \rho(1) = y, \forall t \in [0, 1], W_N^\ell(\rho(t)) < \infty.$$

• Nous posons la conjecture suivante : d'une part pour tout N entier fixé, le nombre des minima locaux de l'énergie électrostatique W_N^ℓ est fini à un

déplacement près et d'autre part la relation d'équivalence R_N^ℓ définie dans l'espace des N -lacets de longueur ℓ par

$$(1.16) \quad \left\{ \begin{array}{l} x R_N^\ell y \iff \\ \text{(il existe un chemin admissible } \rho \text{ entre } x \text{ et } y), (x, y) \in (Y_N^\ell)^2 \end{array} \right.$$

définit l'ensemble Kt_N^ℓ des N -nœuds discrets de longueur ℓ en passant à l'ensemble des paires de lacets, quotienté par la relation d'équivalence R_N^ℓ

$$(1.17) \quad Kt_N^\ell = (Y_N^\ell)^2 / R_N^\ell$$

de sorte qu'un représentant "de forme optimale" de la classe $z \in Kt_N^\ell$ est un minimum local de l'expression (1.14).

- De plus, nous pensons qu'il y a monotonie de l'ensemble Kt^ℓ des nœuds de longueur ℓ , défini par

$$(1.18) \quad Kt^\ell = \{Kt_N^\ell, N \geq 3\}$$

c'est à dire que pour $N_1 \leq N_2$, il existe une injection naturelle de l'ensemble des N_1 -nœuds de longueur ℓ dans celui des N_2 -nœuds de longueur ℓ . A la limite $N \uparrow \infty$, les minima locaux de chaque classe d'équivalence convergent vers la forme optimale universelle d'un nœud donné.

I.4) Approximation numérique.

- Du point de vue du calcul numérique des états définis par la relation (1.14), il s'agit de la minimisation d'une fonction régulière W_N^ℓ définie dans un ensemble Y_N^ℓ contenant N contraintes (voir la relation (1.3)), donc difficile à approcher. Il semble donc naturel de pénaliser le problème d'optimisation (1.14) à l'aide d'un paramètre $\varepsilon > 0$ et de chercher $\bar{x}(\varepsilon) \in X_N$ comme minimum local de la fonction pénalisée suivante :

$$(1.19) \quad W_N^\ell(\varepsilon, x) = W_N(x) + \frac{1}{2\varepsilon} \sum_{j=1}^N \left(|x_j - x_{j-1}| - \frac{\ell}{N} \right)^2.$$

D'un point de vue mécanique, on rajoute au modèle électrostatique décrit au second paragraphe des ressorts entre deux charges consécutives sur le lacet discret. Ces ressorts ont une raideur proportionnelle à $\frac{1}{\varepsilon}$ et sont en équilibre lorsque deux charges consécutives sont distantes de $\frac{\ell}{N}$.

- Le problème (1.19) a le défaut d'ajouter un paramètre de discrétisation au modèle numérique. Par contre, la non convexité de la fonctionnelle est claire et l'emploi de méthodes évolutionnaires (voir par exemple Schoenauer [Sc97]) semble particulièrement adapté à la recherche de tous les minima locaux.

-II- FORME NATURELLE D'UN MONOBRIN D'ADN.

II.1) Introduction.

- Soit L_4 l'alphabet à quatre lettres du code génétique :

$$(2.1) \quad L_4 = \{A, C, T, U\}$$

et M_N l'ensemble des suites finies ordonnées $m = (X_j)_{j=1, \dots, N}$ où chaque lettre X_j appartient à l'alphabet L_4 :

$$(2.2) \quad M_N = (L_4)^N \ni m = (X_1, \dots, X_N); \quad X_j \in L_4, \quad \forall j \in \{1, 2, \dots, N\}.$$

- On cherche la forme naturelle du monobrin m , c'est à dire une application ψ définie sur M et à valeurs dans $(\mathbb{R}^3)^N$ qui associe un point x_j de l'espace tridimensionnel à chaque lettre X_j de m :

$$(2.3) \quad M \ni m \quad \mapsto \quad \psi(m) = (x_1, \dots, x_N) \in (\mathbb{R}^3)^N.$$

- On demande à ψ d'une part (et pour simplifier) d'être compatible avec une distance constante entre les points images : $\psi \in D_a$, c'est à dire

$$(2.4) \quad \exists a > 0, \quad \forall (x_1, \dots, x_N) \in \psi(M), \quad \forall j \in \{1, \dots, N-1\}, \quad |x_{j+1} - x_j| = a$$

et d'autre part de minimiser le potentiel d'interaction du brin, entité que nous définissons au second paragraphe.

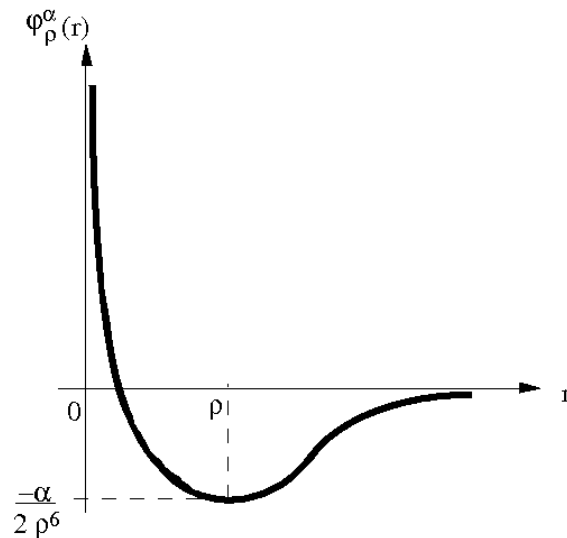


Figure 2 Potentiel de Van der Waals.

II.2) Potentiel d'interaction.

- Le potentiel d'interaction décrit l'attraction privilégiée entre les lettres (A,C) d'une part et (C,U) d'autre part. On se donne par exemple un potentiel d'interaction moléculaire du type "Van der Waals" :

$$(2.5) \quad \varphi_\rho^\alpha(r) = -\frac{\alpha}{r^6} + \frac{\alpha \rho^6}{2 r^{12}}, \quad \alpha > 0, \quad \rho > 0, \quad r > 0$$

qui a une position d'équilibre en $r = \rho$ où la fonction φ_ρ^α est minimale (Figure 2). Du point de vue de la chimie moléculaire de l'acide ribonucléique, cette hypothèse est certainement critiquable (voir par exemple Alberts et al [A195]). Toutefois, le potentiel de Van der Waals est à la fois relativement peu coûteux à calculer et de décroissance à l'infini réaliste. Le changer pour une autre fonction élémentaire d'interaction ne change pas la démarche sur le fond.

- Toute paire $\{X_i, X_j\}$ de la monochaîne m égale soit à la paire $\{A, T\}$ soit à la paire $\{C, U\}$ contribue au potentiel d'interaction. On suppose ici pour simplifier que les interactions $\{A, T\}$ sont décrites par le potentiel de Van der Waals de paramètres ρ et α alors que les interactions $\{C, U\}$ sont paramétrées par la même valeur ρ de la position d'équilibre associée à un autre coefficient positif β . Nous posons

$$(2.6) \quad W^m(\psi) = \sum_{\substack{1 \leq i < j \leq N \\ \{X_i, X_j\} = \{A, T\}}} \varphi_\rho^\alpha(|x_i - x_j|) + \sum_{\substack{1 \leq i < j \leq N \\ \{X_i, X_j\} = \{C, U\}}} \varphi_\rho^\beta(|x_i - x_j|).$$

- Par définition, une forme naturelle du monobrin $m \in M$ est une application $\psi_m \in D_a$ qui minimise le potentiel $W(\psi)$:

$$(2.7) \quad \psi_m \in D_a, \quad W^m(\psi_m) \leq W^m(\psi), \quad \forall \psi \in D_a, \quad m \in M_N.$$

Il est utile de comparer les distances a et ρ de ce modèle géométrique ainsi que la valeur relative des forces α et β des deux potentiels d'interaction pour assurer un maximum de vraisemblance à cette modélisation géométrique. Le calcul d'un tel état de minimum ψ_m est bien entendu délicat pour des brins très longs où l'entier N est grand.

II.3) Pénalisation.

- Un autre modèle consiste à lever la contrainte (2.4) en écrivant que la distance $|x_{j+1} - x_j|$ entre deux lettres X_j et X_{j+1} consécutives correspond à une distance a d'équilibre d'un potentiel moléculaire φ_a^γ de la forme (2.5). On remplace alors la contrainte $\psi \in D_a$ par un terme supplémentaire de pénalisation W_a qui est un potentiel d'interaction de Van der Waals liant deux éléments consécutifs du brin pour pouvoir satisfaire de façon plus faible la contrainte de distance sur la position des lettres du monobrin m :

$$(2.8) \quad W_a(\psi) = \sum_{1 \leq j \leq N-1} \varphi_a^\gamma(|x_{j+1} - x_j|).$$

On pose ensuite

$$(2.9) \quad W_a^m(\psi) = W^m(\psi) + W_a(\psi), \quad m \in M, \quad \psi \in (\mathbb{R}^3)^N.$$

- Une forme d'équilibre du monobrin $m \in M_N$ est alors une solution ψ_m du problème d'optimisation sans contrainte suivant

$$(2.10) \quad \psi_m \in (\mathbb{R}^3)^N, \quad W_a^m(\psi_m) \leq W_a^m(\psi), \quad \forall \psi \in (\mathbb{R}^3)^N$$

et ce problème de minimisation peut aussi être posé sous la forme de la recherche d'un minimum local qui donne d'autres formes possibles, éventuellement instables vis à vis des perturbations dans le temps, d'un monobrin. On cherche alors, pour $m \in M_N$ fixé, une solution ψ_m du problème suivant :

$$(2.11) \quad \left\{ \begin{array}{l} m \in M, \quad \psi_m \in (\mathbb{R}^3)^N, \quad \exists \eta > 0, \quad \forall \psi \in (\mathbb{R}^3)^N, \\ (\psi \in (\text{boule de centre } \psi_m \text{ et de rayon } \eta)) \Rightarrow (W_a^m(\psi_m) \leq W_a^m(\psi)). \end{array} \right.$$

• Le problème (2.10) ou (2.11) est beaucoup plus “simple” que le problème (2.7) puisqu'on a éliminé la contrainte (2.4). Toutefois, le problème (2.10) dépend de l'entier N , des cinq paramètres réels $(a, \rho, \alpha, \beta, \gamma)$ et de la mono chaîne $m \in M$, c'est à dire de la suite de lettres de l'alphabet L_4 . Notons qu'en pratique, l'entier N vaut plusieurs millions (voir par exemple [Wa95]), ce qui conduit à un problème d'optimisation (2.10)-(2.11) de taille quasi-incompatible avec les moyens de calcul automatique actuellement disponibles.

-III- Remerciements.

Quelques discussions très libres avec Jean-Louis Loday en 1997 ont permis de proposer ces modèles discrets pour aborder des problèmes bien loin de mes préoccupations initiales.

-IV- Références bibliographiques.

- [Ad94] C.C. Adams. *The knot book*, Freeman, New York, 1994.
- [Al95] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, J.D. Watson. *Biologie moléculaire de la cellule*, troisième édition, Flammarion, Paris, 1995.
- [Ja62] J.D. Jackson. *Classical electrodynamics*, J. Wiley, New York, 1962.
- [Ka96] V. Katritch, J. Bednar, D. Michoud, R.G.Scharein, J. Dubochet, A Sta-siak. Geometry and physics of knots, *Nature*, vol. 84, p. 142-145, nov. 1996.
- [Sc97] M. Schoenauer. *Méthodes évolutionnaires*, Ecole Polytechnique, 1997.
- [Wa95] M.S. Waterman. *Introduction to computational biology. Maps, sequences and genomes*, Chapman & Hall, London, 1995.