

Acta Cryst. (1990). **A46**, 623–624

Obtention de l'enveloppe moléculaire des protéines par la déconvolution de la fonction de Patterson de ces protéines natives. Par GEORGES SÉBASTIEN CHIGLIEN, *Unité d'Immunologie Structurale, Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris CEDEX 15, France*

(Reçu le 31 août 1988, accepté le 29 janvier 1990)

Abstract

A process, based on the superposition method, is described to delimit the molecular envelope of a protein by deconvolution of the Patterson function. The object is to obtain a preliminary set of phases from native structure-factor amplitudes with only native intensity data. The method has been tested with data from two immunoglobulin Fab fragments, Fab NEW and Fab R19.9. Several zones of resolution were explored.

Nous nous proposons de montrer que l'on peut obtenir l'enveloppe moléculaire des protéines par la déconvolution de la fonction de Patterson de ces seules protéines natives, donc en ne tenant compte que de leurs intensités diffractées. Nous entendons par enveloppe moléculaire l'ensemble des zones de densités dont les valeurs sont plus élevées et réparties d'une façon plus homogène que dans celles représentant le solvant. Elles doivent recouvrir avec une précision globale suffisante, reflétée par une figure de mérite, le domaine de l'espace où se développe effectivement la molécule. Quatre étapes sont nécessaires.

(1) *Méthode des superpositions*

Communément, lorsque l'on considère une structure comme une distribution de pics de Dirac, la fonction de Patterson s'écrit

$$P(\mathbf{u}) = V^{-1} \sum_i \sum_j Z_i Z_j \delta[\mathbf{u} - (\mathbf{r}_i - \mathbf{r}_j)]$$

avec les coordonnées exprimées en fractions de bords de maille; \mathbf{u} le vecteur position du point courant dans la maille Patterson; V le volume de la maille; $\mathbf{r}_i, \mathbf{r}_j$ les vecteurs positions des atomes courants i et j de la structure; Z_i, Z_j le nombre d'électrons des atomes i et j . Chaque pic de la fonction de Patterson représente donc un vecteur interatomique. Une translation de l'origine de la fonction de Patterson égale à l'un de ces vecteurs $\mathbf{t}_{mn} = \mathbf{r}_m - \mathbf{r}_n$ conduit à une nouvelle fonction $P(\mathbf{u} - \mathbf{t}_{mn})$ dont la superposition avec $P(\mathbf{u})$

$$F_{mn}(\mathbf{u}) = P(\mathbf{u}) + P(\mathbf{u} - \mathbf{t}_{mn})$$

ne laisse en coïncidence que les pics dessinant une figure de la structure et son inverse par rapport à un point situé au milieu de \mathbf{t}_{mn} . Dans le cas d'une structure non centrosymétrique le choix d'un autre vecteur interatomique $\mathbf{t}_{pq} = \mathbf{r}_p - \mathbf{r}_q$ et le réitération avec celui-ci des mêmes opérations que prédominamment produit une deuxième figure de la structure accompagnée d'un inverse par rapport à un point situé, cette fois, au milieu de \mathbf{t}_{pq}

$$F_{pq}(\mathbf{u}) = P(\mathbf{u}) + P(\mathbf{u} - \mathbf{t}_{pq}).$$

Ainsi sur la superposition des deux résultats précédents

$$F(\mathbf{u}) = F_{mn}(\mathbf{u}) + F_{pq}(\mathbf{u})$$

les pics en coïncidence dessinent la structure seule (Patterson, 1934, 1935; Wrinch, 1939; Buerger, 1950). Avec les protéines la résolution atomique n'est pas atteinte et, de plus, elles comportent un très grand nombre d'atomes. Par conséquent les pics de la fonction de Patterson se recouvrent systématiquement entre eux. Mais la densité moléculaire forme elle-même une figure continue que l'on peut espérer différencier des zones de solvant.

(2) *Ajustement de l'origine*

Le résultat de la première phase doit être visualisé sous forme de projections. Elles montrent une symétrie éventuellement tronquée. Ce qui en subsiste indique qu'il faut translater la densité obtenue de façon à faire coïncider son origine avec celle de la maille. Quand aucune symétrie n'apparaît il suffit de remplacer chacun des vecteurs de composantes x, y, z utilisés dans la méthode des superpositions par un cône de vecteurs. Ce cône est constitué de tous les vecteurs dont les composantes sont limitées par les bornes $x \pm dx, y \pm dy, z \pm dz$. L'opération de translation-superposition est répétée pour chacun des vecteurs du cône et la superposition commune de cet ensemble de superpositions est uniquement conservée. La multiplication des superpositions provoque un appauvrissement de la densité de déconvolution, seules les zones de très bon recouvrement subsistant, ce qui permet de rendre visibles les restes de symétrie. La translation de l'origine ainsi déterminée est alors appliquée au résultat de déconvolution effectuée sans utilisation des cônes de vecteurs.

(3) *Restauration de la symétrie du groupe d'espace à chaque fois que la symétrie de la densité obtenue n'est que partielle*

Soient respectivement g une opération quelconque de symétrie du groupe d'espace G ; \mathbf{r} le vecteur position du point P courant de la maille M de la densité de déconvolution; $g\mathbf{r}$ le nouveau vecteur position après application de l'opération g à \mathbf{r} ; $\rho(\mathbf{r})$ et $\rho(g\mathbf{r})$ les valeurs des densités aux extrémités de \mathbf{r} et de $g\mathbf{r}$ avant resymétrisation; $\rho_s(\mathbf{r})$ la valeur de la densité à l'extrémité de \mathbf{r} après resymétrisation.

On prendra

$$\forall g \in G \text{ et } \forall P \in M \\ \rho_s(\mathbf{r}) = \max[\rho(\mathbf{r}), \rho(g\mathbf{r})].$$

(4) *Définition d'une figure de mérite afin d'évaluer la validité de chaque solution*

Sont calculés le coefficient de reliabilité

$$R = \sum (\text{FO} - \text{FD} \cdot \text{SC})^2 / \sum \text{FO}^2$$

où

$$\text{SC} = \sum \text{FO} \cdot \text{FD} / \sum \text{FD}^2,$$

Tableau 1. *Figures de mérite*

Molécule	Résolution (Å)	R (%)	FCOR1 (%)	FCOR2 (%)	Nombre de réflexions
Fab NEW	30	4,00	97,98	96,20	12
	9	34,37	81,01	83,90	430
	4,5	34,85	80,72	58,75	2738
Fab R19.9	9	34,57	80,89	75,37	377
	4,5	38,22	78,60	51,68	3098

SC étant le facteur d'échelle qui minimise R , ainsi que deux facteurs de corrélation

$$FCOR1 = \frac{\sum FO.FD}{(\sum FO^2 \sum FD^2)^{1/2}}$$

$$FCOR2 = \frac{\sum (FO - \overline{FO})(FD - \overline{FD})}{[\sum (FO - \overline{FO})^2 \sum (FD - \overline{FD})^2]^{-1/2}}$$

où FO représente les facteurs de structure observés, FD représente les facteurs de structure calculés à partir de la densité de déconvolution.

Nous avons testé cette méthode avec deux molécules de Fab (fraction antigen binding) différentes. D'une part la molécule de Fab NEW (Saul, Amzel & Poljak, 1978) dont seules subsistent les positions atomiques de la molécule

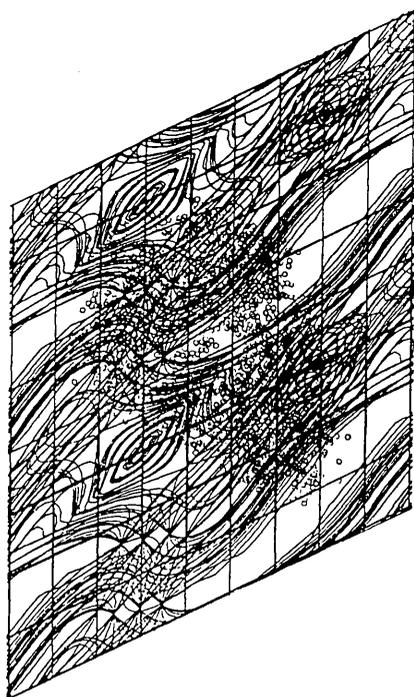


Fig. 1. Fab NEW. Etude à 30 Å de résolution. Superposition de la densité de déconvolution et des positions atomiques publiées. Projection le long de l'axe y . L'axe x est vertical et l'axe z horizontal. Limites en fractions de bords de maille: $x = -0,625$ à $0,375$; $y = -0,250$ à $0,750$; $z = -0,375$ à $0,750$.

résolue à 2 Å. Par conséquent, nous avons travaillé sur les intensités calculées à partir de ces coordonnées. D'autre part le Fab R19.9 (Mariuzza *et al.*, 1984) en cours de résolution au laboratoire et pour lequel étaient disponibles les intensités mesurées jusqu'à 4,5 Å de résolution. Les groupes d'espace $C121$ du Fab NEW et $P12_1$ (*International Tables for Crystallography*, 1983) du Fab R19.9 n'étant pas centrosymétriques chaque essai nécessite de prendre un couple de vecteurs. Avec le R19.9 les cônes de vecteurs ont dû être employés. Pour chacune de ces molécules et dans chaque domaine de résolution exploré (30, 9 et 4,5 Å) plusieurs déconvolutions ont été effectuées. Les meilleurs résultats obtenus sont consignés dans le Tableau 1. On voit que les figures de mérite se dégradent progressivement avec l'augmentation de la résolution. La Fig. 1 concerne le Fab NEW à 30 Å de résolution et résume le bon accord obtenu. Dans le cas du Fab R19.9 les conditions sont moins favorables. La symétrie du groupe d'espace est plus basse de sorte que les cartes de densités, tant celle provenant de la déconvolution que celle issue du phasage MIR (multiple isomorphous replacement) présentent un aspect nettement moins caractéristique. D'une façon générale nous pensons qu'il serait hasardeux d'essayer d'interpréter la densité de déconvolution sans un affinement préalable des phases. Dans les circonstances où l'on ne dispose que des intensités issues du seul cristal d'une protéine native une telle méthode serait susceptible de fournir le point de départ nécessaire à la résolution de la structure.

L'auteur remercie ses collègues de l'Unité d'Immunologie Structurale et ceux de l'Unité d'Informatique Scientifique de l'Institut Pasteur pour l'aide amicale qu'ils lui ont apportée.

Références

- BUERGER, M. J. (1950). *Acta Cryst.* **3**, 87-97.
International Tables for Crystallography (1983). Tome A. Dordrecht: Reidel. (Distributeur actuel Kluwer Academic Publishers, Dordrecht.)
MARIUZZA, R. A., AMIT, A., BOULOT, G., SALUDJIAN, P., SAUL, F., TOUGARD, P., POLJAK, R. J., CONGER, J., LAMOYI, F. & NISSONOFF, A. (1984). *J. Biol. Chem.* **259**, 5954-5958.
PATTERSON, A. L. (1934). *Phys. Rev.* **46**, 372.
PATTERSON, A. L. (1935). *Z. Kristallogr.* **90**, 517.
SAUL, F. A., AMZEL, M. L. & POLJAK, R. J. (1978). *J. Biol. Chem.* **253**, 585-597.
WRINCH, D. M. (1939). *Philos. Mag.* **27**, 98.