Biochemistry (Moscow), Vol. 69, No. 7, 2004, pp. 742-747. Translated from Biokhimiya, Vol. 69, No. 7, 2004, pp. 912-918.

Originally published in Biochemistry (Moscow) On-Line Papers in Press, as Manuscript BM03-246, March 28, 2004.

How Proteins Pack into Crystals: Nuclei Achieve Translation Symmetries by Growing

Dan Feng* and Zong-Hao Zeng

Center of Molecular Biology, Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Chaoyang District, Beijing 100101, China; E-mail: fengdan_bj@hotmail.com

Received October 1, 2003
Revision received November 5, 2003

Abstract—How protein molecules pack into a crystal remains problematic. Packing units are direct materials for packing into crystals. The group generator method is introduced for automatically identifying the packing unit. By introducing deviations into the nucleation stage of crystallization, we proved that these deviations diminish in further packing. This process illustrates how translation symmetries are achieved by the growing of nuclei. Two effects, the size effect and the close up effect, are found to behave differently in this process.

Key words: crystal packing, frustration, packing unit, nucleation, group generator method

The structural characterization of protein–protein interfaces has been studied by many authors [1-15]. They have used many properties, including hydrophobicity, shape analysis, residue preferences, and number of hydrogen bonds, to characterize a protein–protein interface.

In these works, many efforts were made with regard to discriminating between crystal and true oligomeric contacts. Crystal oligomers refer in particular to those oligomeric contacts that can exist only in a crystal but not under physiological condition, such as dimers in the porcine adenylate kinase rhombohedral crystal [16]. In this dimer, the active sites of two monomers are crossed and locked in the interface, and they would not have enzyme activity if they existed in physiological solutions.

We suggest, and other research groups also put forward the same point of view [4], that crystal and true oligomer have a common feature that both of them can exist in advance in pre-crystallization solutions. Results from solution experiments provide powerful support for this argument. The abovementioned dimer in porcine adenylate kinase rhombohedral crystal can be induced by high salt condition, although it cannot exist in physiological solution. Dimer of pancreatic ribonuclease was supposed to be an intermediate of crystallization in high salt condition [3], and this supposition was verified by subsequent experimental spectral data [17]. Dynamic light scattering proved that canavalin formed trimers before crystallization [18]. Introduction of the disulfide cross-links was used to obtain crystals in a different space group of a T4 lysozyme mutant that could not be crystallized previously [19]. The results suggest that the formation of the lysozyme dimer is a critical intermediate in the formation of more than one crystal form.

A schematic picture can be given to packing, or to a crystallization pathway: polypeptide chains aggregate in pre-crystallization solutions and then assemblies pack into crystals. To stress the feature with respect to crystal packing, in this work we call assemblies in pre-crystallization solution as packing units. The interactions between packing units, not within them, are the major forces leading to crystal packing.

Locally preferred interactions between two packing units often have slight deviations from that in the final crystal. Only when frustration is successfully avoided, the crystal grows [20]. But, even in the simplest model system such as an equal-size hard sphere system, frustration exists, let alone in systems of protein molecules, which are much more complex both geometrically and physicochemically. How are the frustrations avoided and which factors influence the long-range order dominating over the local preferred packing? By introducing deviations into the nucleation stage of crystal packing, we proved that these deviations diminish in further packing. This

* To whom correspondence should be addressed.

0006-2979/04/6907-0742 ©2004 MAIK "Nauka/Interperiodica"

process illustrates how translation symmetries are achieved by the growing of nuclei. Two effects, the size effect and the close up effect, are found to behave differently in this process.

GROUP GENERATOR METHOD

Our goal is to identify the cluster in which all chains contact with each other more closely than with the rest in the crystal. We defined the cluster as the packing unit of this crystal. There are essentially two keys in recognizing it. The first is how to evaluate the extent of contact. To do this, the size of the solvent accessible surface area (ASA) [21] buried in the interaction has been commonly used. The ASA describes the extent to which a protein can form contacts with water. Contact area of two molecules is the area of the protein surface that becomes buried in contacts between molecules, evaluated as the sum of the solvent accessible surface area of two partners minus that of the pair. ASA of a molecule was calculated by the SURFACE program (CCP4) (UK).

The second key is concerned with the algorithm. Before we continue to explain this problem, we make an insight into the various way that protein molecules associate together. According to interface type in the final assembly, assemblies are divided into two kinds: **A** type and **B** type. In **A** type assemblies, any one of the interfaces is generated by two monomers. In other words, all interfaces are monomer–monomer interface. **B** type assemblies can be seen as complex **A** type, in which monomers are substituted by a cluster of chains. That is to say, interfaces in **B** type assemblies are generated not only by monomer–monomer interaction, but also by cluster–cluster interaction. In addition, assemblies can be divided into **a** type and **b** type according to the way of topological binding. In **a** type assemblies, contents of the asymmetric unit (ASU) are in contact with all other symmetry related members of the final assembly, otherwise we call them **b** type. The right half of Fig. 1 shows schematic diagrams of the assembly structures in the left half. Small circles represent monomers, while clusters are represented by large ellipses. We do not show **A** type assemblies in this figure. But the contents in each cluster can be seen as simple **A** type oligomer. From top to bottom, the insulin R_6 hexamer is a **Ba** type assembly constituted by three **Aa** type dimers; the hepatitis virus δ antigen decamer is **Bb** type and also is constituted by four **Aa** type dimers; the peptidase dodecamer is **Ba** type constituted by two **Ab** type hexamers.

Henrick et al. [22] used the progressive addition method to find potential quaternary assemblies in a crystal. Chain selection to the assembly is based on the number of inter-chain contacts found and the number of residues in each chain. As the authors said themselves, this method was recursive in detecting quaternary structures of **b** type assemblies. In addition, because this method did not consider the interactions between clusters, it was not so efficient for finding **B** type assemblies when only a small interface existed between monomers.

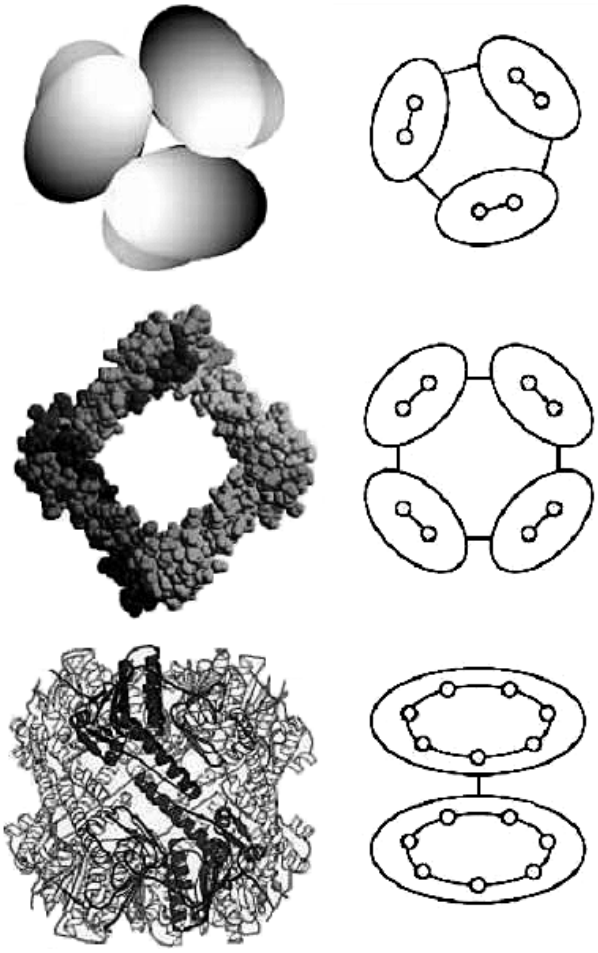

Fig. 1. Examples of packing units with complex binding diagrams. From top to bottom: the insulin R_6 hexamer [26, 27], hepatitis virus δ antigen [28], and a peptidase, respectively [29]. The ellipsoids are used to show the insulin B-chain helices in insulin R_6 hexamer. Rasmol [30] and Molscript [31] were used to plot the other two oligomers.

In the present work, we introduce the group generator method to discriminate the packing unit from the crystal. The strategy of this method is to find a packing unit generating group (PUGG). Like the space group can produce all the coordinates in the crystal when acting on ASU, PUGG produces the coordinates of a packing unit when acting on ASU.

We take the ASU containing a monomeric chain or an oligomer as an example. Collect those symmetry operations that can produce all the direct contact neighbors of the ASU into a set of P. Then the PUGG is generated from P by taking in all possible products of operations in P. We can prove that PUGG contains only and all opera-

tions that can generate all symmetry ASUs in the final packing unit (proof omitted).

Obviously, the group generator method avoids recursive calculation in finding **b** type assemblies. In addition, a cluster can be seen as a new single chain to begin a new cycle calculation in identifying **B** type assemblies.

TWO EFFECTS IN CRYSTAL PACKING

The most powerful function of the packing unit is that contacts in crystal can easily be divided into two parts, those inside the packing unit and those between. Interactions inside the packing unit happen in the stage of pre-crystallization solution, while crystal packing is mainly governed by interactions between packing units. In the following, we begin with packing units of two typical crystals to analyze two effects that can weaken or eliminate frustrations in the nucleation stage.

Size effect. Frustrations in the very early stage of packing can gradually be corrected with the enlargement of the system size. Take the pig insulin crystal (PDB ID 4INS) [23] as an example. Use contact area as a criterion to evaluate the strength of contacts. Hexamers are identified as packing units from the crystal. These hexamers are connected along their 3-fold symmetry axis, with a contact area of 1763 Å^2 (interface A shown in Fig. 2). Other contacts between hexamers are produced by primitive translations (in the R lattice), and each produce 640 Å^2 contact area (interface B). It is most probable that hexamers pack along their 3-fold axis first, since interface A is much larger. Further packing of these line-aggregates in a crystal requires that no rotation exists inside the line, to ensure each hexamer can contact with other hexamers from neighboring line-aggregates with the same conformation.

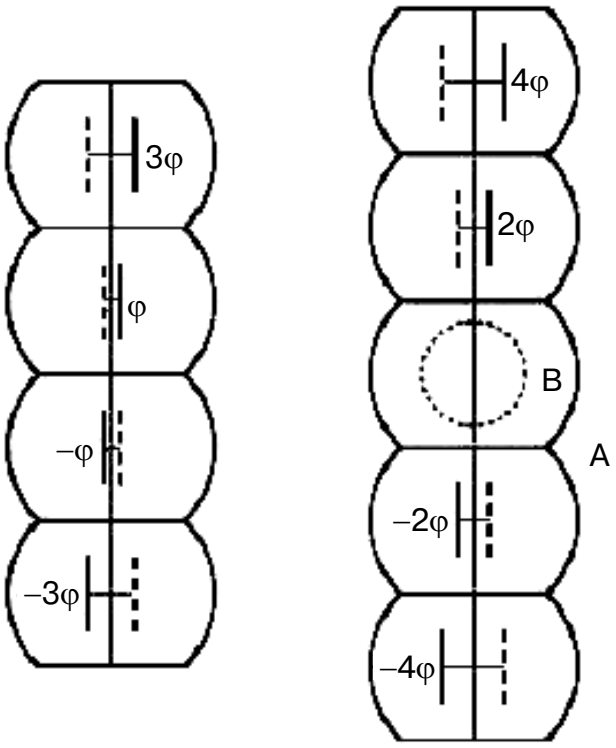

Fig. 2. Packing of two line-aggregates of insulin hexamer with even (left) or odd (right) number of hexamers. The two line-aggregates are parallel to, one below and another above, the paper. The heavy lines denote the centers of contact site B on the two line-aggregates (one solid and another dashed).

Suppose there is a preferred relative rotation, φ, when two hexamers aggregate along their 3-fold axis. The binding potential of interface A, F_A, varies with the relative rotation angle φ:

$$F_A = F_A^0 - k_A(\varphi - \varphi_0)^2/2, \quad (1)$$

where k_A is the elastic constant of interface A and F_A^0 is the strength of the locally preferred binding at interface A (when $\varphi = \varphi_0$). The non-zero value of φ will introduce a miss-match ψ on each interface B when two line-aggregations pack together. The binding potential of interface B, F_B, varies with angle ψ:

$$F_B = F_B^0 - k_B\psi^2/2, \quad (2)$$

where k_B is the elastic constant of interface B and F_B^0 is the strength of the locally preferred binding at interface B.

For two line-aggregates each with n hexamers, there are $2(n-1)$ interfaces of A and n interfaces of B. The miss-match ψ can be list as φ, ±3φ, ±5φ, ..., $\pm(2m-1)\varphi$, for $n = 2m$, or 0φ, ±2φ, ±4φ, ..., $\pm 2m\varphi$, for $n = 2m + 1$ (Fig. 2). Therefore, the total packing strength F can be expressed as:

$$F = 2(n-1)F_A^0 - (n-1)k_A(\varphi - \varphi_0)^2 + \\ + nF_B^0 - n(n^2-1)k_B\varphi^2/6. \quad (3)$$

To get this result formulas on sums of square integers have been used. Under equilibrium, $F/\varphi = 0$, which leads to the equilibrium value of φ_e as:

$$\varphi_e = \varphi_0/(1 + \xi), \quad (4)$$

with

$$\xi = n(n+1)(k_B/k_A)/6. \quad (5)$$

As the number n increases, φ decreases to zero with the rate of $1/n^2$. That means that deviations in the very early stage of crystal packing are gradually corrected with extending of the line-aggregates. When φ is below the thermodynamic uncertainty, the packing between the two

line-aggregates becomes the same as that in the final crystal. This model reflects the size effect in avoiding deviations.

Close up effect. The bungarotoxin crystal (PDB ID 1KBA) [24] represents another model of avoiding deviations. We identified a dodecamer as this crystal's packing unit. The binding diagram is cyclic and the dodecamer has symmetry of 6-fold rotation axis. In the shape of a hexagon, each dodecamer has six contact sites with other dodecamers, with a contact area of 2335 Å^2 much larger than the contact areas of about 140 Å^2 produced by packing along the 6-fold axis. This leads to plane-aggregates in a hexagonal lattice in two-dimensional space.

The contact sites between the dodecamers are at their six edges. Suppose the most preferred conformation for binding of two dodecamers is not necessarily in a plane. The two planes of the two dodecamers may make an angle δ_0. But, this angle δ_0 must be sufficiently small. Otherwise, the system may pack into a polyhedron. Now, suppose that in the most preferred conformation with an angle δ_0 the binding strength of the two dodecamers is F^0 and in plane conformation the strength is $F = F^0 - k\delta_0^2/2$, with k as the corresponding elastic constant. Then, for a system consisting of three dodecamers, there can only form two contacts with total strength $2F^0$ if all of them are in the most preferred conformation, while they can form three contacts with total strength $3(F^0 - k\delta_0^2/2)$ if three dodecamers are in the same plane. Therefore the condition for the plane conformation to be selected is:

$$F^0 > 3k\delta_0^2/2. \quad (6)$$

More generally, we consider the case of "closed shells": one central dodecamer surrounded by six dodecamers in the first shell, 12 dodecamers in the second shell, etc. (Fig. 3). If there are n shells around a central dodecamer, then altogether there are $3n(n + 1) + 1$ dodecamers. In plane conformation $9n^2 + 3n$ contacts can form, while at most $3n(n + 1)$ contacts can exist in the locally preferred conformation. Therefore, the condition for the system to select the plane conformation is:

$$F^0 > \xi k\delta_0^2/2, \quad (7)$$

with

$$\xi = (3n + 1)/2n. \quad (8)$$

When n is very large, $\xi \to 1.5$.

When all dodecamers locate in the same plane, there are not cleavages between hexagon edges. This is the close up effect. The close up effect makes the system either in a plane, or not, by condition (7). The elementary unit for the close up effect in this example is three dodecamers bound through three contacts. When more dodecamers join the aggregation, the condition (7) becomes weaker with the decreasing of ξ. That means the aggregate is increasingly easier to be in a plane with the enlargement of the system. The decrease in ξ is due to the reduction of average free binding sites at the edge of the aggregate, i.e., a kind of "surface effect" in 2-dimensions. But ξ in condition (8) approaches a finite value, which means that the system always needs to make selection on being in plane or out of plane. If the system selected the out of plane aggregation and failed to form a closed form with all the binding sites being satisfied, then it would most probably precipitate.

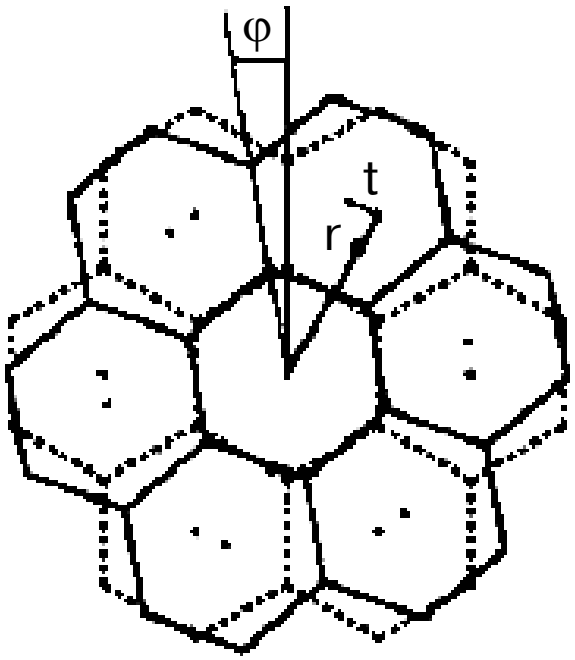

Fig. 3. Two plane-aggregates, respectively in solid and dashed lines, pack one above the other. Only the center hexagon and those in the first "closed shell" are shown. The mismatch t caused by a relative rotation φ tends to cumulate if more shells exist.

DISCUSSION

We have managed to analyze how proteins pack into crystals. Our efforts are focused on answering the question of how the locally preferred protein–protein interactions, which are usually not prepared for crystallization by evolution, lead to long-range order.

Based on current solution experimental evidence, we suggest that there are pre-existing assemblies in a pre-crystallization solution. These assemblies are direct materials for further packing into crystal.

Packing units in crystals are potential quaternary structure of protein. They are related to, but not necessarily identical to true protein oligomers or assemblies investigated in solution or *in vivo*. The differences lie not

only in the degrees of oligomerization but also on the conditions for their survival. We do not focus on discriminating true oligomerization states from packing units in this work. Instead, we pay more attention to the fact that they are direct materials to pack into crystals. Contacts responsible for crystal packing can be extracted and analyzed separately.

The group generator method is especially developed to identify the packing unit from the crystal. The basis of this method is to find the packing unit generating group, which is a subgroup of the space group. The packing unit generating group contains all and only symmetry operations that can generate those symmetries of ASUs in the final assemblies when acting on the ASU. The group generator method has advantages in automatic searching for packing units of a large number of crystals. It overcomes recursive calculation and is more efficient in finding oligomers with a complex topological structure.

It is natural to suppose that, at the very early stage of packing, e.g., in the nucleation process, there are deviations in the preferred interface from that in final crystal and these preferred interactions lead to aggregates in lower-dimensional space. Therefore, these low-dimensional aggregates, if kept unchanged, could not pack into crystals. Only when further packing successfully corrects the low dimensional packing into that suitable for crystallization, the frustration can be avoided. This analysis follows that of Pauling [25] on proteins. He suggested that proteins developed two complementary regions when the solution condition was changed. If the two regions led to helical aggregates, in general, they would not involve a rational number of monomers per turn and would accordingly not be well suited to the formation of crystals. Interactions between these helical aggregates might deform the helices slightly into configurations with a rational number of molecules per turn, resulting in the formation of crystals. When different interfaces with various strengths, and accordingly various privileges, intend to repeat themselves, the geometry of the three-dimensional space makes restrictions on their repeatability.

By introducing deviations from the crystal packing in the lower stage aggregates, and analyzing how these deviations vanish while the packing proceeds further, two effects were revealed to behave differently at the early time of crystallization. First, the size effect tells us that, as the size of the system increases, the frustration, i.e., the difference between the most preferred local conformation and the local conformation existing in the final crystal, diminishes. The frustration decreases with the rate of $1/n^2$, where n is the size of the system counted by number of molecules. When packing proceeds from one-dimension or two-dimension to higher dimensional space, the size effect helps to overcome frustrations. Size effect also acts in the second example. Plane-aggregates will pack up parallel in the next stage. We also can prove that slight rotation between two neighbor layers can be weakened with the enlargement of the system (Fig. 3, proof omitted). This effect decreases the frustration gradually, and formally, never makes it zero. When the frustration is decreased below thermodynamic uncertainty, a kind of order–disorder transition is expected to happen.

When a molecule has more than two binding sites, as in the case of the hexagonal dodecamers in the second example, the close up effect can happen at each of the binding sites and bring the packing into two-dimensional space by forming plane-aggregates. The system has to select to either be symmetric (in plane) or be asymmetric (out of plane). Although the condition for an aggregate in a plane is increasingly easier to be satisfied with the increase in the size of the system, the system cannot avoid making such a selection.

Compare the two effects. The size effect is like a tune-up switch. It minimizes the errors but never eliminates them completely. The close up effect is like an on-off switch. It ensures the crystal being in no-error status increasingly easier, but with possible stop in any time.

This work is supported by the National Natural Science Foundation of China (grant No. 30080004).

REFERENCES

1. Janin, J., Miller, S., and Chothia, C. (1988) *J. Mol. Biol.*, **204**, 155-164.
2. Janin, J., and Chothia, C. (1990) *J. Biol. Chem.*, **265**, 16027-16030.
3. Crosio, M. P., Janin, J., and Jullien, M. (1992) *J. Mol. Biol.*, **228**, 243-251.
4. Janin, J., and Rodier, F. (1995) *Proteins*, **23**, 580-587.
5. Jones, S., and Thornton, J. M. (1995) *Progr. Biophys. Biol.*, **63**, 31-65.
6. Jones, S., and Thornton, J. M. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 13-20.
7. Jones, S., and Thornton, J. M. (1997) *J. Mol. Biol.*, **272**, 121-132.
8. Jones, S., and Thornton, J. M. (1997) *J. Mol. Biol.*, **272**, 133-143.
9. Lijnzaad, P., and Argos, P. (1997) *Proteins*, **28**, 333-343.
10. Tsai, C. J., Lin, S. L., Wolfson, H. J., and Nussinov, R. (1997) *Protein Sci.*, **6**, 63-64.
11. Carugo, O., and Argos, P. (1997) *Protein Sci.*, **6**, 2261-2263.
12. Janin, J. (1997) *Nat. Struct. Biol.*, **4**, 973-974.
13. Ponstingl, H., Henrick, K., and Thornton, J. M. (2000) *Proteins*, **41**, 47-57.
14. Jones, S., Marin, A., and Thornton, J. M. (2000) *Protein Eng.*, **13**, 77-82.
15. Elcock, A. H., and McCammon, J. A. (2001) *Proc. Natl. Acad. Sci. USA*, **98**, 2990-2994.
16. Dreusicke, D., Karplus, P. A., and Schulz, G. E. (1988) *J. Mol. Biol.*, **199**, 359-371.
17. Jullien, M., Crosio, M. P., Baudet-Nessler, B., Merola, F., and Brochon, J. C. (1994) *Acta Crystallogr.*, **D50**, 398-403.
18. Kadima, W., McPherson, A., Dunn, M. F., and Jurnak, F. A. (1990) *Biophys. J.*, **57**, 125-132.

19. Heinz, D. W., and Matthews, B. W. (1994) *Protein Eng.*, **7**, 301-307.
20. Sadoc, J.-F., and Mosseri, R. (1999) *Geometrical Frustration*, Cambridge University Press, Cambridge.
21. Lee, B., and Richards, F. (1971) *J. Mol. Biol.*, **55**, 379-400.
22. Henrick, K., and Thornton, J. M. (1998) *Trends Biochem. Sci.*, **23**, 358-361.
23. Baker, P. J., Sawa, Y., Shibata, H., Sedelnikova, S. E., and Rice, D. W. (1998) *Nat. Struct. Biol.*, **5**, 561-567.
24. Dewan, J. C., Grant, G. A., and Sacchettini, J. C. (1994) *Biochemistry*, **33**, 13147-13154.
25. Pauling, L. (1952) *Discus. Faraday Soc.*, **13**, 170-176.
26. Brader, M. L., and Dunn, M. F. (1991) *Trends Biochem. Sci.*, **16**, 341-345.
27. Ding, J., Wan, Z., Chang, W. R., and Liang, D. C. (1996) *Sci. China C Life Sci.*, **39**, 144-153.
28. Zuccola, H. J., Rozzelle, J. E., Lemon, S. M., Erickson, B. W., and Hogle, J. M. (1998) *Structure*, **6**, 821-830.
29. Wang, J., Hartling, J. A., and Flanagan, J. M. (1997) *Cell*, **91**, 447-456.
30. Sayle, R. A., and Milner-White, E. J. (1995) *Trends Biochem. Sci.*, **20**, 374.
31. Kraulis, P. J. (1991) *J. Appl. Crystal.*, **24**, 946-950.