FEBS Letters 525 (2002) 176–178 FEBS 26255

Correspondence

Protein structure: experimental and theoretical aspects

Joerg Harms[a], Frank Schluenzen[a], Raz Zarivach[b], Anat Bashan[b], Heike Bartels[a], Ilana Agmon[b], Ada Yonath[a,b,*]

First published online 12 July 2002

Using self developed software, intended to assess the quality of experimentally determined three-dimensional structures of proteins, Bujnicki et al. spotted irregularities in the folds of a few proteins in 1kc9, the original deposition of the coordinates of the large ribosomal subunit from *Deinococcus radiodurans* (D50S). Independently, we, the crystallographers that determined this structure at 3.1 Å resolution [1], discovered about a week after the release of 1kc9 that this entry was damaged and over 80% of its content (~65 000 coordinates) was ruined. Within 10 days we fixed the coordinates file and re-deposited the coordinates (1kpj).

Re-deposition of coordinates to the data bank is rather common in ribosomal crystallography. Most recently Ramakrishnan's group deposited a new entry of the coordinates of the small ribosomal subunit from *Thermus thermophilus* (T30S) (1j5e). That refers to their original, almost 2 years old, article [2]. Preliminary analysis of 1j5e indicates new placements of several RNA bases, including those involved in decoding that were originally flipped out as well as minor shifts in a few proteins, most noticeable in S17.

The original entry of the coordinates of the large subunit from *Haloarcula marismortui* (H50S), 1ffk, was deposited in August 2000 [3]. A second deposition (1jj2) appeared in August 2001. Although the H50S structure was determined at rather high resolution, 2.4 Å, the two H50S depositions show several significant differences. Nevertheless, Bujnicki et al. did not detect any problems in either deposition. On the contrary, they claim that 'the scores of the polypeptide chains in both entries (1jj2 and 1ffk) were generally high, indicating satisfactory quality of both H50S structures'.

As an example of the problematic proteins in H50S we chose protein L15e. The two depositions of this protein show only 74% sequence identity, and the second contains two additional residues. Within a total of 173 residues, there are over two dozen substitutions. Among them are 10 arginines that were replaced by dramatically different amino acids, such as glycine, alanine, glutamic and aspartic acids, serine, tryptophan, tyrosine, valine and proline. Additional examples include K-Y, Q-Y, A-M, E-P, F-P, T-P, H-F, S-W, C-A etc. Nevertheless, as mentioned above, Bujnicki's method found these two extremely different depositions to be satisfactory.

Bujnicki et al. used 1kpj to create theoretical models for the D50S proteins and to compare them with our experimental results. This comparison revealed deviations between the models and the structures of four (out of 31) proteins. One of the four suspected proteins is L16. This is one of the most important proteins in the large ribosomal subunit, since it interacts with the A-site tRNA hence may moderate the binding of the tRNA molecule to the ribosome. In both H50S depositions no protein was identified as being homologous to the eubacterial L16. However, once we determined the structure of D50S the similarity of L16 to L10e in their structures (Fig. 1) and in the relative locations in D50S and in H50S became evident [1]. Consequently, Steitz and Moore, the senior authors of the H50S depositions, reconfirmed these similarities (in a talk given at the Ribosome meeting, January 2002; and in a plenary lecture at the International Biophysics Society Meeting in Buenos Aires, April 2002).

L10e is one of the proteins that show significant inconsistencies between the two H50S entries, 1ffk and 1jj2. This protein has 167 residues in the recent entry, but only 156 in the original one. Among the additional adjacent 11 residues, 10 are of an unknown sequence. These were inserted in the middle of the protein, around position 100, between two residues that were connected in the original deposition. Additional modifications include non-conventional substitutions, such as K-G, K-Y, K-D (twice), E-F, N-G, N-A, N-V, N-L, I-D, P-S, V-S, I-S, M-H and Y-F.

Ironically, the method of Bujnicki et al. found both of these very different depositions (1jj2 and 1ffk) to be satisfactory, and at the same time it identified so-called 'errors' in L16 of D50S, the protein that led to the identification of its counterpart in H50S (L10e) as having a prokaryotic origin, rather than eukaryotic, as originally assigned.

L15 is a problematic protein. This observation was made by Bujnicki et al. as well as by us. Parts of this protein are indeed less well resolved in our electron density map. To partially overcome this problem, we attempted its tracing based on homologies, using several algorithms. However, none of them led to fully satisfactory parameters, as judged in terms of short contacts, packing geometry, and contacts with the neighboring rRNA or solvent molecules. Unfortunately, Bujnicki et al. did not release their model; hence we cannot comment about its fit to the electron density map as well as to the environment of this protein.

In fact, Bujnicki et al. have not provided any description of how they tested various crucial parameters, in terms of energy and stereochemistry, nor did they address the chemical or biological meaning of their models. The ribosome is extremely crowded, and it is conceivable that Bujnicki's theoretical models are placed in otherwise occupied space. On the other hand, the proteins must interact with their neighborhood. In addition, Bujnicki et al. did not optimize their models in terms of secondary structure hydrogen bonds, or those created between the ribosome and the solvent or the metal ions. They also did not supply crucial definitions concerning the reliability or the resolution of their method.

A few additional proteins of D50S suffer from shortcomings similar to those of L15. Among them is the flexible protein L27. Bujnicki et al. did not mention this protein, presumably because there is no H50S counterpart to L27, hence a structural homologue that could supply the basis for Bujnicki's prediction is not available. Using known structures for predicting three-dimensional structures of proteins may some-

PII: S0014-5793(02)02960-5

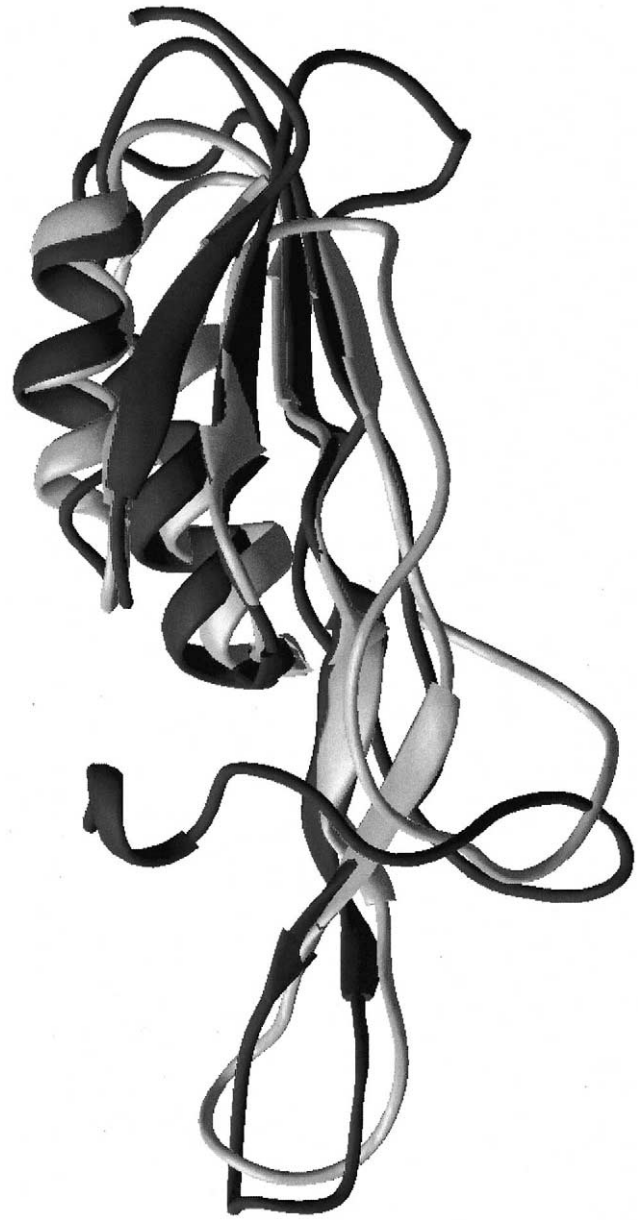

Fig. 1. Superposition of the H50S protein L10e (black) on its D50S counterpart L16 (gray).

times yield approximately correct results, but it may also be very dangerous. In cases when a single structure is the sole source for the prediction, even reasonable homology does not assure correct prediction, since the predicted structure is bound to carry all the shortcomings of the one used as the 'structural base'.

The remaining suspected proteins in Bujnicki's list are L2 and L24. We see no obvious problems with protein L2, although it is one of the longest chains of the large subunit. Protein L24 was, until recently, incomplete. In both D50S depositions, as well as in many additional PDB entries, segments of several proteins are missing since they are not visible in the electron density maps. Supplement S1 [1] shows the exact numbers of traced vs. expected amino acids for each of the D50S proteins. Discrepancies between the actual traced and the expected DNA sequences may stem from structure flexibility, ambiguous regions in the electron density map due to proximity to ordered solvent or metal ions, or simply because some codons are not being expressed.

Difficulties in tracing portions of protein chains are not unique to D50S. Significant portions of proteins were not traced in the 3–3.2 Å electron density maps of the small subunit, T30S [2,4,5]. Furthermore, in contrast to the general belief, this problem cannot be simply correlated with the resolution, as it was found to be more prominent at higher resolution. Thus, four proteins could not be seen in the 2.4 Å map of the large ribosomal subunit from H50S [3] and do not even appear in 1ffk (3.75 proteins in 1jj2). In addition, large portions of other H50S proteins were not traced. An interesting example is the entire domain of protein L5 that creates an intersubunit bridge, which is missing in both H50S depositions, 1ffk and 1jj2.

Protein L24 was problematic during the initial tracing. Only recently, when we extended the resolution of the D50S structure to 3 Å (PDB deposition 1lnr), could we trace L24 with confidence. Interestingly, the 3 Å structure of L24 differs somewhat from the original one, but shows only partial consistency with Bujnicki's suggested alterations.

Bujnicki et al. claim that structure factors should provide 'the ultimate testing of their models vis-a-vis the experimental data'. It is not clear how Bujnicki et al. intend to use the structure factors for calculating the reliability of their predictions. Inspection of electron density maps should provide the required information. However, we doubt whether such studies fit the aims of bioinformatics, namely to confirm or raise doubts about experimental results within the shortest possible time-frame.

We therefore assume that Bujnicki et al. plan to use the experimental observations for checking the reliability of their predictions, exploiting measures such as the *R* factor. However, according to our experience, such criteria are bound to fail in such tasks due to the low sensitivity of the factor to changes in single proteins. In addition, 'blind' refinement of biological structures (i.e. refinement of *R* or similar factors without referring to maps or chemical properties) was proven extremely dangerous even for very small proteins that diffract to very high resolution. Three decades ago Scheraga and colleagues showed that the least-squares refined 1.55 Å structure of Rubridoxin [6], that was determined with the impressive *R* factor of $<20\%$, contained wrong bond lengths, bad dihedral angles and unacceptable inter-atomic distances [7]. When energy refined, the *R* factor was raised to 37%, but the chemical nature of the molecule made sense. Consequently, restraining and constraining the refinement became a general procedure (e.g. [8]), and currently the correctness of the stereochemistry and similar parameters are routinely being verified when coordinates are deposited in the data bank. We assume that Bujnicki et al. aim at developing a useful tool to be incorporated in the verifications required by the PDB during deposition. We believe that in the future they may achieve this aim. However, in view of its failure to recognize the problems in the H50S depositions, and the overestimation of the irregularities in the structures of the D50S proteins, it is clear that in its current state, Bujnicki's method is far from being suitable for this aim and needs to be significantly improved. Meanwhile it is important to stress that the studies of Bujnicki et al. show, at most, discrepancies between theoretical predictions and experimental observations. So far there is no reason to refer to these discrepancies as 'errors'.

References

[1] Harms, J., Schluenzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., Bartels, H., Franceschi, F. and Yonath, A. (2001) Cell 107, 679–688.
[2] Wimberly, B.T., Brodersen, D.E., Clemons Jr., W.M., Morgan-Warren, R.J., Carter, A.P., Vonrhein, C., Hartsch, T. and Ramakrishnan, V. (2000) Nature 407, 327–339.
[3] Ban, N., Nissen, P., Hansen, J., Moore, P.B. and Steitz, T.A. (2000) Science 289, 905–920.
[4] Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F. and Yonath, A. (2000) Cell 102, 615–623.
[5] Pioletti, M., Schlunzen, F., Harms, J., Zarivach, R., Gluhmann, M., Avila, H., Bashan, A., Bartels, H., Auerbach, T., Jacobi, C., Hartsch, T., Yonath, A. and Franceschi, F. (2001) EMBO J. 20, 1829–1839.
[6] Watenpaugh, K.D., Sieker, L.C., Herriott, J.R. and Jensen, L.H. (1973) Acta Crystallogr. B29, 94–156.
[7] Rasse, D., Warme, P.K. and Scheraga, H.A. (1974) Proc. Natl. Acad. Sci. USA 71, 3736–3740.
[8] Susmman, J.L., Holbbrook, S.R., Church, G.M. and Kim, S.H. (1977) Acta Crystallogr. A33, 800–804.

*Corresponding author.
E-mail address: yonath@mpgars.desy.de (A. Yonath).

[a] *Max-Planck-Research Unit for Ribosomal Structure, Hamburg, Germany*
[b] *Department of Structural Biology, Weizmann Institute, Rehovot, Israel*

PII: S0014-5793(02)02960-5