FEBS Letters 525 (2002) 174–175 FEBS 26254

Correspondence

Errors in the *D. radiodurans* large ribosomal subunit structure detected by protein fold-recognition and structure validation tools

Janusz M. Bujnicki*, Marcin Feder, Leszek Rychlewski, Daniel Fischer

First published online 20 June 2002

Using protein fold-recognition methods and the acclaimed tool for protein structure quality verification Verify3D, we have identified potential errors in the *Deinococcus radiodurans* large ribosomal subunit structure (D50S; Protein Data Bank entry 1kc9) and proposed theoretical models of seven protein structures (L13, L14, L15, L16, L18, L23, and L24; deposited in the Protein Data Bank as 1gs2). The authors of the D50S structure have been immediately notified and provided with the names of the potentially mistraced chains, they have confirmed that 1kc9 contained a number of shifts, and deposited a replacement entry (1kpj) without explicitly utilizing information from our models. In the replacement entry, predicted shifts in proteins L13, L14, L18, and L23 have been corrected in agreement with the theoretical models, thereby validating our methodology. However, our analysis strongly suggests that proteins L2, L15, L16, and L24 in the 1kpj entry (chains B, J, K, and S) still contain errors and may require further revision, even though they have already been retraced compared to 1kc9. Unfortunately, the authors of the D50S structure have not provided structure factors, which impeded ultimate testing of our models vis-à-vis the experimental data.

Here we present our findings of errors persisting in the 'corrected' *Deinococcus radiodurans* large ribosomal subunit (LSU) structure (D50S; 1kpj in Protein Data Bank (PDB) [1]), using as example the L24 protein (chain S). The evidence of a poor sequence-structure fit in the D50S structure is obtained from two independent sources. The first source is evolutionary information gathered from homologous sequences and the distantly related structure of *Haloarcula marismortui* LSU structure (1jj2 [2]) (Fig. 1a). The multiple sequence alignment highlights the conserved patterns between 1kpjS and 1jj2S (and other bacterial and archaeal L24 proteins). A similar sequence-structure alignment has been reported by fold-recognition methods that optimally aligned the *D. radiodurans* L24 protein sequence to the sequence and structure of its homolog from *H. marismortui* (comprehensive fold-recognition results are available from the protein structure prediction Meta Server [3] at the URL http://bioinfo.pl/meta/target.pl?id=5053). However, a structural superposition generates a shifted alignment between 1kpjS and 1jj2S in the regions $18–30^{(+1\ aa)}/31–68^{(+2\ aa)}/69–115^{(+3\ aa)}$ (*D. radiodurans* protein L24 residue numbers). We used Modeller [4] to build a model of the *D. radiodurans* L24 protein (1gs2S), based on the original 1kc9S coordinates but using different sequence-to-structure assignment, consistent with the sequence conservation pattern in the L24 protein family.

The second source of evidence is provided by the established method Verify3D [5] for structure quality verification (Fig. 1b). Since the ribosomal protein data in the D50S structure included only their C-α coordinates, the full-atom representation (required by Verify3D) was reconstructed using Maxsprout (http://www.ebi.ac.uk/dali/maxsprout/ [6]). As a control, we applied Maxsprout to the earlier release of the *H. marismortui* LSU structure (1ffk [7]), which includes only C-α coordinates of the ribosomal proteins. The full-atom representations of all ribosomal proteins in the entries 1jj2 (refined using experimental data) and 1ffk (reconstructed using Maxsprout) were very similar and their Verify3D scores were not significantly different (data not shown). Moreover, despite the presence of various structural idiosyncrasies in ribosomal proteins [7], the scores of polypeptide chains in the entries 1jj2 and 1ffk were generally high, indicating satisfactory quality of the H50S structure.

However, the overall Verify3D score of the full-atom reconstruction of 1kpjS turned out to be very low (0.143), with the lowest dips corresponding to our predicted shifts. This is usually indicative of serious problems in the structure. In contrast, the overall score of our theoretical model 1gs2S is acceptable (0.271), with no dips below 0.1. These two sources of information are independent of the experimental data, and as such, they can provide valuable indications of potential shifts in the tracing of the amino acid sequence into a particularly difficult electron-density map. Similar cogent evidence from sequence analysis and structure evaluation is obtained for proteins L2, L15, and L16 (data not shown), suggesting that, in order to obtain a more accurate understanding of the molecular details of the ribosome, a further revision of the D50S structure may be required. In addition to the theoretical models of seven D50S proteins deposited in PDB (1gs2), full-atom models of all D50S proteins are available from J.M.B. (e-mail: iamb@genesilico.pl).

While this article was processed for publication, our report of errors in the very first version of the *D. radiodurans* LSU structure (1kc9), which was submitted before the replacement entry (1kpj) was made available, was accepted for publication as 'Fold-recognition detects an error in the Protein Data Bank', J.M. Bujnicki, L. Rychlewski and D. Fischer, Bioinformatics (in press).

Note added in proof

While reviewing the proofs of our paper, we discovered that the replacement entry 1kpj has now been replaced by a third pdb entry (1lnr). In 1lnr, chains D, H, O and S have been significantly revised (amino acid identities have been shifted). A major part of the S chain has been retraced essentially in agreement with our model presented in this paper.

References

[1] Harms, J., Schluenzen, F., Zarivach, R., Bashan, A., Gat, S.,

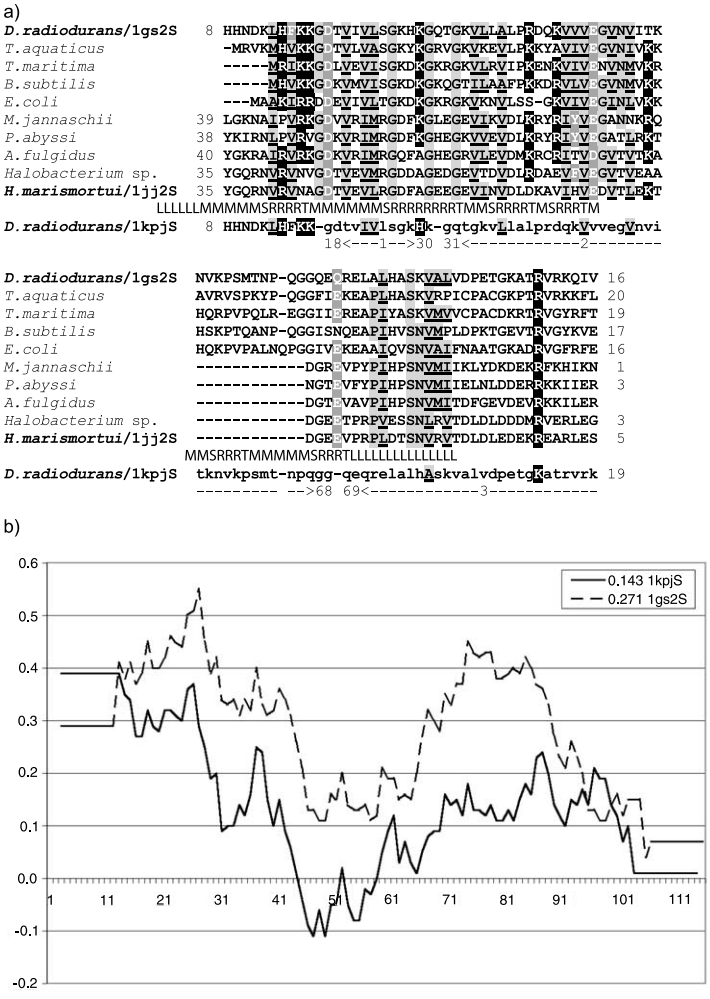

Fig. 1. a: Sequence alignment reveals inconsistency of 1kpjS with evolutionary conservation among the L24 proteins. Residues are shaded according to their physicochemical properties. The conserved patterns are highlighted. The number of omitted terminal residues is indicated. The secondary structure of 1jj2S is shown as tubes and arrows. The structural alignment of 1kpjS and 1j22S is shown at the bottom, with the three predicted shifts labeled 1, 2, and 3. b: Structure evaluation with Verify3D reveals that poor regions (score <0.1) in 1kpjS are relieved in the theoretical model 1gs2S.

Agmon, I., Bartels, H., Franceschi, F. and Yonath, A. (2001) Cell 107, 679–688.
[2] Klein, D.J., Schmeing, T.M., Moore, P.B. and Steitz, T.A. (2001) EMBO J. 20, 4214–4221.
[3] Bujnicki, J.M., Elofsson, A., Fischer, D. and Rychlewski, L. (2001) Bioinformatics 17, 750–751.
[4] Sali, A. and Blundell, T.L. (1993) J. Mol. Biol. 234, 779–815.
[5] Luthy, R., Bowie, J.U. and Eisenberg, D. (1992) Nature 356, 83–85.
[6] Holm, L. and Sander, C. (1991) J. Mol. Biol. 218, 183–194.
[7] Ban, N., Nissen, P., Hansen, J., Moore, P.B. and Steitz, T.A. (2000) Science 289, 905–920.

*Corresponding author. Fax: (48)-22-668 5288.
E-mail: iamb@genesilico.pl (J.M. Bujnicki).

Bioinformatics Laboratory, International Institute of Molecular and Cell Biology, ul. ks. Trojdena 4, 02-109 Warsaw, Poland

PII: S0014-5793(02)02959-9