Biochemical and Biophysical Research Communications 431 (2013) 65–69

Contents lists available at SciVerse ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Small-angle X-ray scattering constraints and local geometry like secondary structures can construct a coarse-grained protein model at amino acid residue resolution

Yasumasa Morimoto, Takashi Nakagawa, Masaki Kojima *

School of Life Sciences, Tokyo University of Pharmacy and Life Sciences, Horinouchi, Hachioji, Tokyo 192-0392, Japan

ARTICLE INFO

Article history:
Received 18 December 2012
Available online 3 January 2013

Keywords:
Coarse-grained model
Restrained molecular dynamics
Secondary structure
Small-angle X-ray scattering

ABSTRACT

We have developed a new methodology that determines protein structures using small-angle X-ray scattering (SAXS) data. The current bottlenecks in determining the protein structures require a new strategy using the simple design of an experiment, and SAXS is suitable for this purpose in spite of its low information content. First we demonstrated that SAXS constraints work additively to NMR-derived information in calculating structures. Next, structure calculations for nine proteins taking different folds were performed using the SAXS constraints combined with the NMR-derived distance restraints for local geometry such as secondary structures or those for tertiary structure. The results show that the SAXS constraints complemented the tertiary-structural information for all the proteins, and that accuracy of the structures thus obtained with SAXS constraints and local geometrical restraints ranged from 1.85 to 4.33 Å. Based on these results, we were able to construct a coarse-grained protein model at amino acid residue resolution.

1. Introduction

The three-dimensional structure of proteins provides critical information about their biological function. At present a large number of protein structures have been determined by X-ray crystallography and NMR. These experimental methods, however, have intrinsic barriers during sample preparation such as crystallization and isotope labeling, as well as the methodological problems such as unambiguous signal assignments and phase determination. On the other hand, the pure theoretical prediction of protein structures from their amino acid sequence remains imperfect in spite of the recent developments of such approaches [1,2], especially in case where no homologous proteins are available. These situations require the establishment of a new strategy for determining the protein structure using the simple design of an experiment combined with a knowledge-based theoretical approach.

Small-angle X-ray scattering (SAXS) is one of the experimental methods suitable for this purpose [3]. It provides the structural information of molecules in solution. Since the amount of information from SAXS is less than that from X-ray crystallography or NMR mainly due to the fluctuation and the random orientation of molecules in solution, SAXS provides only knowledge about overall shape and size of the molecules. Despite its fallbacks, however, SAXS has two major advantages over the above-mentioned problems in structure analysis. The first is the ease in preparing samples and adjusting measurement conditions. The second is the simplicity in data treatments that allows us to get the output information directly after the experiment. Recent advances in *ab initio* modeling enable us to construct a coarse-grained structure as an ensemble of small beads with SAXS data alone [4,5].

We previously developed a new program *SAXS_MD*, based on the restrained molecular dynamics (MD) algorithm with SAXS constraints [6]. It rationally calculates the constraint force applied to each atom, and effectively changes its position in the molecule so that the theoretical scattering intensity curve calculated from the current model matches well with the experimental SAXS data. Given the appropriate initial structure, this approach can successfully construct the valid three-dimensional structure. Since the restrained MD is also used in refining the structure obtained from X-ray crystallography or NMR [7], it is possible to combine SAXS constraints with various types of structural information in the calculation. Actually, the combination of SAXS and NMR data in the restrained MD calculation has already been reported [8,9]. These studies, however, aim to refine the multidomain protein structure more accurately by adding SAXS information to the NMR data.

In this study, we first characterized the structural information that SAXS provides in terms of the complementarity to the secondary or tertiary structure of the proteins. As described above, the structural information from SAXS is little, and it is difficult to

---

* Corresponding author. Fax: +81 42 676 5498.
E-mail address: mkojima@toyaku.ac.jp (M. Kojima).

0006-291X/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.bbrc.2012.12.091

reconstruct the protein structure from the SAXS data alone. Then we evaluated what kind of structural information is needed to construct a protein structure in addition to the SAXS constraints. We used NMR-derived distance restraints for complementing SAXS constraints, because NMR also provides the structural information in solution and can construct the full structure at atomic resolution by itself [10]. Finally we have succeeded in constructing coarse-grained protein structures at the amino acid residue resolution from the SAXS constraints and the secondary structure of the proteins. We thus propose a new methodology that can construct a coarse-grained protein structure more rapidly than the conventional ways.

2. Materials and methods

2.1. Calculation procedures

In this study, we performed the restrained MD calculation with SAXS constraints and subsets of NMR-derived distance restraints to construct the three-dimensional molecular model, and analyzed the results obtained under various types of structural information. For this purpose the NMR-derived distance restraints should be classified according to whether each of them defines secondary structures or tertiary structure of each protein. We classified these restraints into a subset supporting the local geometry and its complementary subset, because it is difficult to unambiguously determine which restraints can build a secondary or tertiary structure. Restraints belonging to the former subset were defined to involve in the distance between atoms located within four residues along the polypeptide chain, and the other restraints were defined to belong to the latter subset. The restraints about local geometry define local structures of the molecule such as secondary structures or intraresidue conformation, while the other restraints contribute mainly to the formation of a tertiary structure.

All the calculations were carried out by restrained MD program, *EMBOSS* [11]. Since the program was originally designed for NMR-based structure determination, we integrated our previously developed *SAXS_MD* program into *EMBOSS* as a subroutine and performed the calculation. Considering the hydration effect in SAXS data [12], we performed the restrained MD calculation in dimensionless scale [13], which eliminated the uniform expansion of whole molecule as shown in Fig. 1A. For each calculation with the same input, 30 independent runs were repeated from various initial structures with random torsion angles and the same amino acid sequence, and the generated conformation ensemble was averaged in each atomic position by the same program. As a reference structure, we carried out the same calculation with full sets of NMR-derived distance restraints for comparison. The degree of accuracy of the results was evaluated by the average of root mean square deviations (RMSDs) for backbone atoms of each conformer in the obtained ensemble and the reference structure. Fig. 1B shows a flow diagram that depicts all the calculation procedures described above.

2.2. SAXS and NMR data

We used RNase T1 as the first model protein, because its NMR-derived three-dimensional structure was once determined in our laboratory [14] and the SAXS data measured under the same experimental condition were also available [6].

In addition to RNase T1, eight proteins with different folding topologies based on the Structure Classification Of Proteins (SCOP) [15] were selected according to the following requirements:

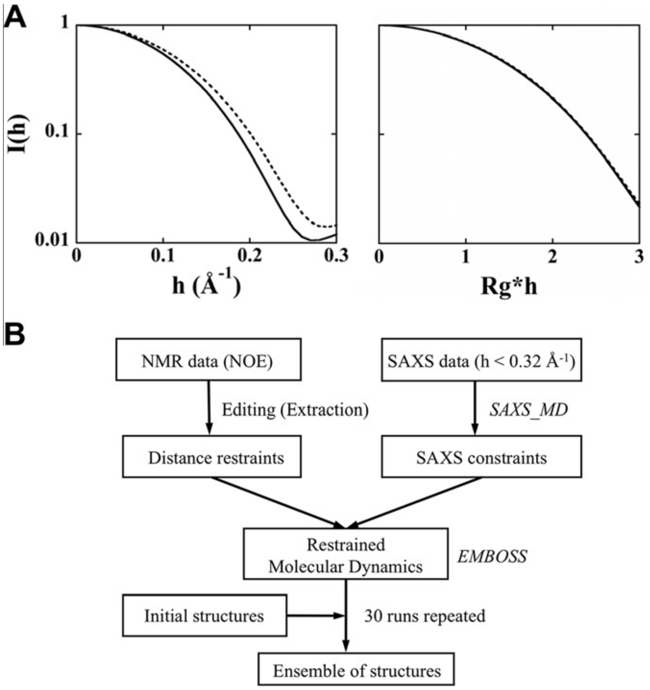

Fig. 1. Calculation procedures with SAXS constraints. (A) Effects of hydration on normal (left) and dimensionless (right) plots. All the SAXS profiles were calculated from the PDB coordinates 1IYY [14] with *CRYSOL* [18] program. Hydration shell was considered by setting the contrast of the solvation shell to 0.03 (default), or not considered by setting the same parameter to 0.00. The resultant SAXS profiles with and without hydration shell are presented as a solid and a dotted line, respectively. (B) Flow diagram of the calculation procedures described in materials and methods section.

(i) The atomic coordinates determined by NMR are available from Protein Data Bank (PDB).
(ii) The NMR-derived distance information obtained experimentally is also available from BioMagResBank (BMRB).
(iii) The same structure as in PDB can be constructed correctly from the NMR-derived structural information in BMRB using the *EMBOSS* program used in this study.

Among eight proteins thus selected, ATC2521 from *Agrobacterium tumefaciens* (PDB ID: 2JQ4) and steril-α-motif of human deleted in liver cancer 2 (2JW2) are classified as all-α proteins, Filamin-B (2DIA) and putative lipoprotein from *Bacillus cereus* (2K5W) as all-β proteins, pyruvate phosphate dikinase (2FM4 [16]) and eukaryotic translation termination factor eRF1 (2HST [17]) as α/β proteins, and Ral guanosine dissociation stimulator (2B3A) and NE1680 from *Nitrosomonas europaea* (2HFQ) as α+β proteins.

For NMR-derived structural information only NOE data were used, because they have direct relationships with interatomic distance and can be replaced by alternative methods in future. Then the restraints were translated into the form of *EMBOSS* input before the calculation. For the eight proteins selected from PDB and BMRB, theoretical SAXS profile was produced by the *CRYSOL* program [18] for default parameter settings in the range of h from 0.00 to 0.32 Å^{-1} where $h = (4\pi/\lambda)\ \sin\theta$, 2θ is a scattering angle and λ is an X-ray wavelength.

The degree of coincidence of two scattering curves was evaluated by the R-factor in the dimensionless scale defined as

$$R = \sum_{R_g h} |I_1(R_g h) - kI_2(R_g h)| \cdot (R_g h)^2 \Big/ \sum_{R_g h} I_1(R_g h) \cdot (R_g h)^2$$

where $I_1(R_gh)$ and $I_2(R_gh)$ are the scattering intensity of the first and the second curve as a function of R_gh, respectively, and k is a scaling factor calculated as

$$k = \sum_{R_gh} I_1(R_gh) \cdot I_2(R_gh) \cdot (R_gh)^2 / \sum_{R_gh} I_2(R_gh)^2 \cdot (R_gh)^2.$$

3. Results and discussion

3.1. Structure calculation of RNase T1

For RNase T1, we first investigated the effect of SAXS constraints to NMR-derived distance restraints in calculating the protein structure. Fig. 2A shows the conformation ensembles calculated with 1076 NMR-derived restraints in the presence and the absence of SAXS constraints. Average structures of both ensembles were almost identical (RMSD = 0.71 Å), and the addition of SAXS constraints improved the convergence between conformers in the ensemble. In addition, *R*-factor of the SAXS pattern calculated from the obtained model against that used as constraints was 6.0% (Fig. 2B), which indicated that the resultant structure with SAXS constraints satisfied well the experimental SAXS profile. These results indicated that the effect of SAXS constraints to NMR-derived restraints was additive rather than conflictive during the calculation.

Next the structure calculations were repeated by gradually reducing the NMR distance information as an input until the resultant structure was wholly collapsed. When 50% of the local geometrical restraints were randomly removed from the original NMR data set (597 restraints), the obtained structure contained no secondary structures and the RMSD to the reference structure was 6.97 Å. We then performed the structure calculations with the same NMR restraints adding the SAXS constraints. The calculated structure, however, was almost the same as that without the SAXS constraints; there were no secondary structures formed and the average RMSD to the reference structure was 6.37 Å (data not shown), which indicated that the SAXS constraints did not complement the local geometry such as secondary structures for RNase T1.

Similarly, when 70% of the tertiary-structural information was randomly removed from the full NMR data set (479 restraints) as an input, the calculated structure was wholly distorted and the RMSD to the reference structure was 4.88 Å (Fig. 2C). We then recalculated the structure with the same NMR restraints in the presence of the SAXS constraints. As a result, all the conformers in the obtained structure ensemble were well superimposed onto the reference structure, and the average RMSD values to the reference was 3.84 Å (Fig. 2C), which indicated that the SAXS constraints partially complemented the tertiary structure for RNase T1.

3.2. Structure calculation of other proteins

To investigate whether the SAXS constraints generally complement the tertiary structure in proteins, we analyzed other eight proteins with different folding topology by the same approach as RNase T1 described above. As shown in Fig. 3, for all the eight model proteins both the degree of convergence of conformers in the structure ensemble and that of accuracy against the reference structure were greatly improved by adding the SAXS constraints to the local geometrical restraints from NMR data. *R*-factors for the obtained structures were 7.9% (2JQ4), 3.0% (2JW2), 9.8% (2DIA), 2.5% (2K5W), 4.6% (2FM4), 4.3% (2HST), 9.8% (2B3A), and

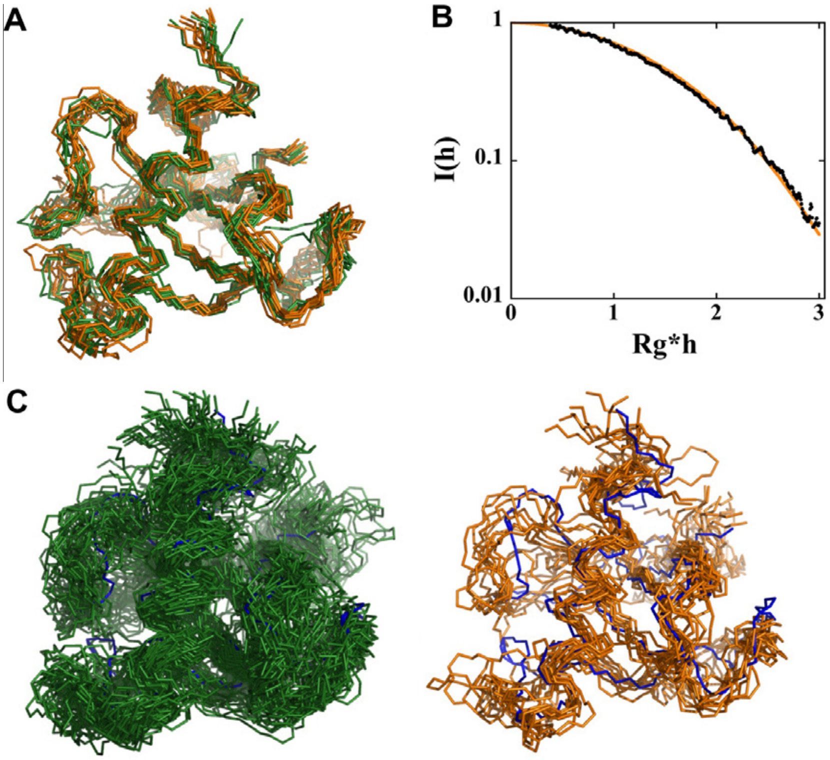

Fig. 2. Structure calculation of RNase T1. (A) Superposition of structures calculated with (orange) and without (green) SAXS constraints as well as full sets of NMR-derived distance restraints. (B) Dimensionless plots of the experimental SAXS profile used as constraints (black) and that calculated from the obtained structure (orange). (C) Superposition of structures calculated with NMR-derived restraints for local geometry in the absence (green) and presence (orange) of SAXS constraints. Reference structure is also presented in blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

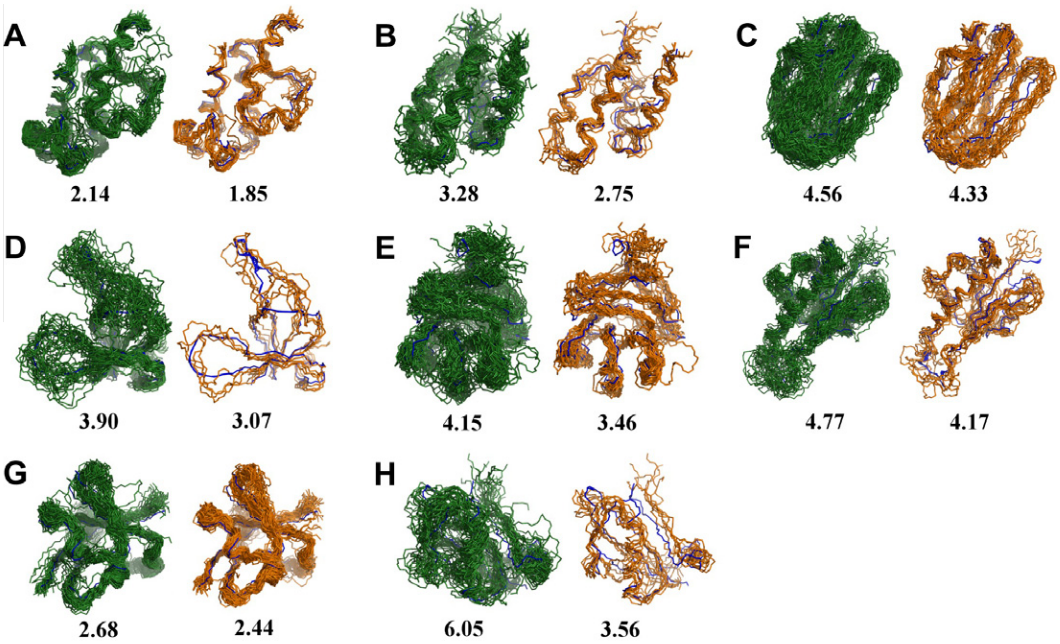

Fig. 3. Calculated structures of eight proteins. Structures calculated with NMR-derived restraints for local geometry in the absence (green) and presence (orange) of SAXS constraints were superimposed along with reference structure (blue). The average RMSD values to the reference structure are also presented in Å. PDB ID of the protein is 2JQ4 (A), 2JW2 (B), 2DIA (C), 2K5W (D), 2FM4 (E), 2HST (F), 2B3A (G) and 2HFQ (H). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

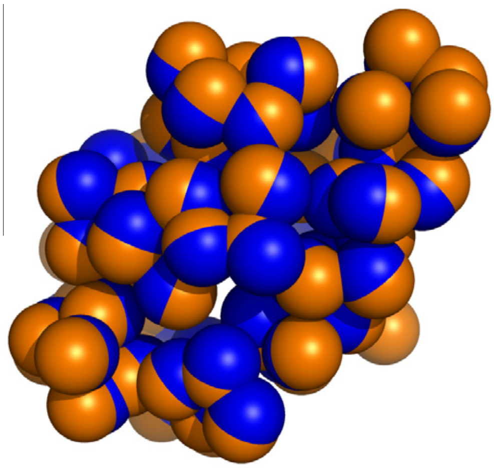

Fig. 4. Coarse-grained structure constructed for 2JQ4. Constructed model with SAXS constraints and local geometrical restraints is presented in orange, and the reference structure in blue. Each residue is illustrated as a sphere with radius of 3.0 Å smaller than its real size on average. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

7.2% (2HFQ). These values indicated that the obtained structures also satisfied well the constraining SAXS profile.

Considering $h = 0.32$ Å^{-1} corresponds to the resolution of about 20 Å, the SAXS constraints complemented the tertiary structure of these proteins, and rebuilt a molecular model up to amino acid residue resolution in the presence of the NMR-derived local geometrical information. These results suggested the possibility to construct the protein structure at residue resolution using the SAXS constraints and the local geometry such as secondary structures. Actually, both the formation of secondary structures and the compaction of the whole molecule are supposed to be essential factors in protein folding [19]. The above findings for the structure calculation are consistent with this supposition.

3.3. Construction of coarse-grained molecular model

Furthermore, we generated a coarse-grained model by placing a bead (small sphere) in the position of the center of each amino acid residue in the molecule. For all the proteins described above, the coarse-grained model from the calculated structures was well superimposed onto that from the reference structure. Fig. 4 shows an example for 2JQ4. It should be noted that in the Figure each residue was represented as a bead with radius of 3.0 Å. Considering that radius of an approximated sphere for amino acid residues ranges from 2.9 Å for Gly to 4.6 Å for Trp [20], it was also confirmed that accuracy of the coarse-grained structure above was around amino acid residue resolution.

In this study, we constructed coarse-grained structures for nine proteins at the residue resolution only with experimental or theoretical SAXS constraints and NMR-derived information about local geometry such as secondary structures. The latter, however, should be replaced with that derived from an alternative approach in future for convenience of the methodology. The structural information about local geometry consisted of restraints for secondary structures (for example, 94% for RNase T1) and those for intraresidue conformation (6% for RNase T1). Since intraresidue restraints have little relationship with the residue-based molecular model, we can focus on the restraints for secondary structures. One of the possible methods that can provide the secondary-structural information instead of NMR is vacuum-ultraviolet circular dichroism (VUVCD) measured below 190 nm. VUVCD spectra provide segment numbers of secondary structures and can predict sequence-based secondary structures combined with neural-network approach [21]. Another candidate to yield the secondary-structural information is wide-angle X-ray scattering (WAXS). Hirai et al. [22] revealed that experimental WAXS curves for seven proteins with different fold can be well reproduced based on Debye equation, and that they are additive in terms of secondary structures. The same approach as VUVCD may be also applicable to WAXS data.

The residue-based structures may be themselves useful to extract pharmacophores in drug discovery, or they will be refined to the atomic resolution by purely theoretical approaches such as rotation of side chains in an exhaustive manner [23].

Acknowledgments

The experimental SAXS profile was used from data measured with permission from the Program Advisory Committee of the Photon Factory (Proposal No. 2003G317). This work was in part supported by Platform for Drug Discovery, Informatics, and Structural Life Science from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

[1] S. Raman, R. Vernon, J. Thompson, M. Tyka, R. Sadreyev, J. Pei, D. Kim, E. Kellogg, F. DiMaio, O. Lange, L. Kinch, W. Sheffler, B.-H. Kim, R. Das, N.V. Grishin, D. Baker, Structure prediction for CASP8 with all-atom refinement using Rosetta, Proteins 77 (2008) 89–99.

[2] K. Lindorff-Larsen, S. Piana, R.O. Dror, D.E. Shaw, How fast-folding proteins fold, Science 334 (2011) 517–520.

[3] M.H.J. Koch, P. Vachette, D.I. Svergun, Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution, Q. Rev. Biophys. 36 (2003) 147–227.

[4] D.I. Svergun, Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing, Biophys. J. 76 (1999) 2879–2886.

[5] D.I. Svergun, M.V. Petoukhov, M.H.J. Koch, Determination of domain structure of proteins from X-ray solution scattering, Biophys. J. 80 (2001) 2946–2953.

[6] M. Kojima, A.A. Timchenko, J. Higo, K. Ito, H. Kihara, K. Takahashi, Structural refinement by restrained molecular-dynamics algorithm with small-angle X-ray scattering constraints for a biomolecule, J. Appl. Crystal. 37 (2004) 103–109.

[7] A.T. Brünger, X-PLOR version 3.1 Manual, Yale University Press, New Haven, CT, 1992.

[8] A.G. Grishaev, J. Wu, J. Trewhella, A. Bax, Refinement of multidomain protein structures by combination of solution small-angle X-ray scattering and NMR data, J. Am. Chem. Soc. 127 (2005) 16621–16628.

[9] F. Gabel, B. Simon, M. Nilges, M. Petoukhov, D. Svergun, M. Sattler, A structure refinement protocol combining NMR residual dipolar couplings and small angle scattering restraints, J. Biomol. NMR 41 (2008) 199–208.

[10] J. Cavanagh, W.J. Fairbrother, A.G. Palmer, N.J. Skelton, Protein NMR Spectroscopy: Principles and Practice, Academic Press, London, 1996.

[11] T. Nakai, A. Kidera, H. Nakamura, Intrinsic nature of the three-dimensional structure of proteins as determined by distance geometry with good sampling properties, J. Biomol. NMR 3 (1993) 19–40.

[12] D.I. Svergun, S. Richard, M.H.J. Koch, Z. Sayers, S. Kuprin, G. Zaccai, Protein hydration in solution: experimental observation by X-ray and neutron scattering, Proc. Natl. Acad. Sci. USA 95 (1998) 2267–2272.

[13] A.A. Timchenko, B.S. Melnik, H. Kihara, K. Kimura, G.V. Semisotnov, GroES co-chaperonin small-angle X-ray scattering study shows ring orifice increase in solution, FEBS Lett. 471 (2000) 211–214.

[14] K. Hatano, M. Kojima, E. Suzuki, M. Tanokura, K. Takahashi, Determination of the NMR structure of Gln25-ribonuclease T1, Biol. Chem. 384 (2003) 1173–1183.

[15] A.G. Murzin, S.E. Brenner, T. Hubbard, C. Chothia, SCOP: a structural classification of proteins database for the investigation of sequences and structures, J. Mol. Biol. 247 (1995) 536–540.

[16] Y. Lin, J.D. Lusin, D. Ye, D. Dunaway-Mariano, J.B. Ames, Examination of the structure, stability, and catalytic potential in the engineered phosphoryl carrier domain of pyruvate phosphate dikinase, Biochemistry 45 (2006) 1702–1711.

[17] E.V. Ivanova, P.M. Kolosov, B. Birdsall, G. Kelly, A. Pastore, L.L. Kisselev, V.I. Polshakov, Eukaryotic class 1 translation termination factor eRF1 – the NMR structure and dynamics of the middle domain involved in triggering ribosome-dependent peptidyl-tRNA hydrolysis, FEBS J. 274 (2007) 4223–4237.

[18] D. Svergun, C. Barberato, M.H.J. Koch, CRYSOL – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates, J. Appl. Crystal. 28 (1995) 768–773.

[19] K.A. Dill, L. MacCallum, The protein-folding problem, 50 years on, Science 338 (2012) 1042–1046.

[20] A.A. Zamyatnin, Amino acid, peptide, and protein volume in solution, Annu. Rev. Biophys. Bioeng. 13 (1984) 145–165.

[21] K. Matsuo, H. Watanabe, K. Gekko, Improved sequence-based prediction of protein secondary structures by combining vacuum-ultraviolet circular dichroism spectroscopy with neural network, Proteins 73 (2008) 104–112.

[22] M. Hirai, M. Koizumi, T. Hayakawa, H. Takahashi, S. Abe, H. Hirai, K. Miura, K. Inoue, Hierarchical map of protein unfolding and refolding at thermal equilibrium revealed by wide-angle X-ray scattering, Biochemistry 43 (2004) 9036–9049.

[23] N. Pokala, T.M. Handel, Energy functions for protein design: adjustment with protein-protein complex affinities, models for the unfolded state, and negative design of solubility and specificity, J. Mol. Biol. 347 (2005) 203–227.