Organic & Biomolecular Chemistry

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/obc

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

Solution structure of a cucurbit[8]uril induced compact supramolecular protein dimer

Dung T. Dang,[a] Ralph P. G. Bosmans,[a] Christian Moitzi,[b] Ilja K. Voets,[c] and Luc Brunsveld*[a,c]

Received (in XXX, XXX) Xth XXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX

DOI: 10.1039/b000000x

Supramolecular assembly of a beta-barrel protein via cucurbit[8]uril results in compact z-shaped protein dimers. SAXS data reveal the formation of a well ordered protein dimer, notwithstanding being connected by a reversible and flexible peptide linker, and highlight the supramolecular induced interplay of the proteins, analogous to covalently linked proteins.

Synthetic protein architectures are at the forefront of research in the chemical sciences.[1] Obtaining control over protein assemblies by synthetic and molecular approaches are attractive entries for investigating and modulating the proteins at hand and for novel biomaterials and diverse biomedical applications.[2] Apart from covalent scaffolds, supramolecular chemistry is an ideal entry to generate these protein architectures and provide reversible control over their formation.[3] Two complementary synthetic supramolecular elements can for example mediate protein biomolecule assembly,[4] and crystal structures of proteins bound to supramolecular host molecules have provided information on such complexes at the atomic level.[5] Notwithstanding all these beautiful examples, detailed insight into the exact molecular shape of the resulting protein architectures in solution has only recently been obtained for a selected type of complex,[6] but in most cases is absent. Such insights are however highly needed to understand the modulatory effect of the supramolecular elements on protein structure and activity.[5,7]

Cucurbit[8]uril is a small concave molecule that binds twofold to the short tripeptide phenylalanine-glycine-glycine (FGG) with high association constant (K_{ter} = 1.5 x 10^{11} M^{-2}).[8] We have previously shown, using spectroscopic techniques, that proteins featuring this small, genetically encoded N-terminal FGG peptide motif can be brought to dimerize[9] or tetramerize[10] upon the simple addition of cucurbit[8]uril. Here, we now the describe the detailed molecular and structural analysis of a cucurbit[8]uril induced protein dimer. Using Dynamic light scattering (DLS) and solution-based small angle X-ray scattering (SAXS), the first molecular structure of a supramolecular induced protein dimer in solution is established, revealing a highly compact, z-shaped, supramolecular protein dimer and providing molecular insights in the supramolecular mediated protein assembly process.

Dynamic light scattering (DLS) experiments on an expressed and purified FGG-tagged monomeric yellow fluorescent protein (FGG-mYFP) showed the hydrodynamic radius, R_H, of this protein to be 2.7 nm (Fig. 1a). This theoretically corresponds to a spherical particle / protein of around 34 kDa and fits nicely to the calculated mass of the FGG-mYFP, of 28 kDa. Upon addition of cucurbit[8]uril to the protein, the R_H increased to 3.2 nm, fully in accord with the expected $2^{1/3}$ increase of R_H upon doubling of the hydrodynamic volume. This R_H corresponds to a theoretical protein of around 51 kDa (Fig. 1a). (The mass calculations are based on a spherical model, possibly not exactly matching the actual shape of the protein dimer.) This size nicely corresponds to that of the calculated mass, 56 kDa, of a cucurbit[8]uril dimerized FGG-mYFP protein and indicates a compact structure. Addition of an excess of memantine, a highly potent competitor for cucurbit[8]uril binding,[11] reverts the supramolecular assembly back to the protein monomeric state. Control experiments on a MGG-mYFP protein, not capable of binding to cucurbit[8]uril,[9] show that cucurbit[8]uril addition has no significant effect on the measured R_H of this protein (Fig. 1b).

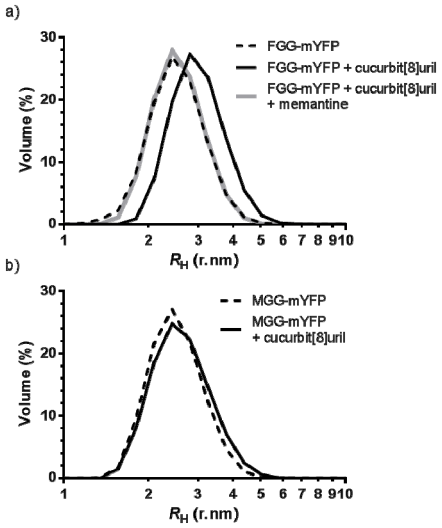

Fig. 1 Dynamic light scattering data for a) FGG-mYFP and b) MGG-mYFP at 40 µM in absence (dotted black line) and presence of cucurbit[8]uril at 20 µM (straight black line) and after addition of memantine (40 µM), a strong cucurbit[8]uril binder (grey line).

Small angle X-ray scattering (SAXS) has emerged as a useful technique to study biological macromolecules in solution, and can provide molecular information on the formation and shape of supramolecular protein complexes in solution.[12-14] Therefore, SAXS studies were performed on the single FGG-mYFP protein

and the FGG-mYFP·cucurbit[8]uril complex (Fig. 2a). The scattering patterns superimpose in the intermediate to high q-regime, $q \geq 0.8$ nm^{-1}, indicating that the structure of the individual domains is very similar in all the different samples. The SAXS patterns of the supramolecular complex at two different concentrations are virtually indistinguishable, indicating that interparticle interference effects are negligible.

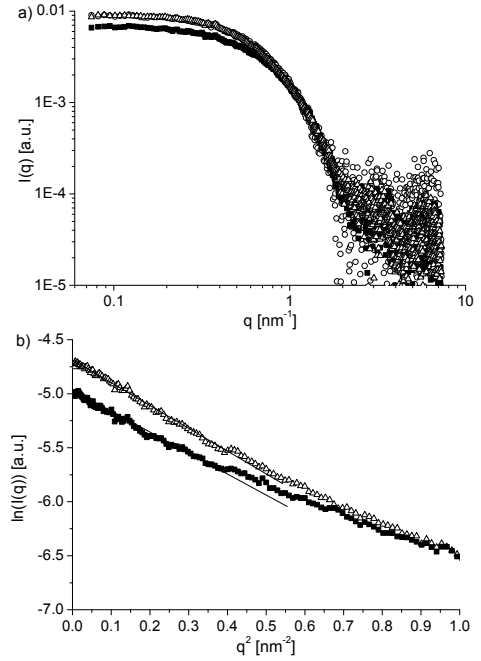

Fig. 2 Background subtracted and concentration normalized small angle X-ray scattering curves of monomeric FGG-mYFP (910 µM, filled squares) and cucurbit[8]uril-induced FGG-mYFP dimer at two different concentrations (857 µM FGG-mYFP / 428 µM cucurbit[8]uril, open triangles; 214 µM FGG-mYFP / 107 µM cucurbit[8]uril, open circles) in 10 mM sodium phosphate buffer at pH 7. b) Corresponding guinier representation of the SAXS patterns.

The radius of gyration (R_G) and forward scattering intensity $I(0)$ were determined from a Guinier analysis from linear fits to the data in the q-range $0.1 \leq q \leq 0.65$ nm^{-1} (Fig. 2b), which is valid in the limit of small scattering vectors ($qR_G < 1$), according to:

$$I(q) = I(0)e^{-q^2 R_G{}^2/3}$$

The results are listed in Table 1. The R_G, and the forward scattering intensity normalized by concentration, $I(0)/c$, are larger for FGG-mYFP in the presence of cucurbit[8]uril. Experimentally a ratio of 1:1.6 is found for the $I(0)/c$ of the samples with and without cucurbit[8]uril, which is larger than the theoretical ratio of 1:2 expected for cucurbit[8]uril induced dimerization of monomeric FGG-mYFP. Furthermore, we obtain $R_G = 2.2$ nm for the sample without curcubit[8]uril, which is slightly larger than $R_G = 1.8$ nm calculated for the YFP parent protein without the linker and FGG motif, based on the high resolution crystal structure 1YFP.[15] We attribute these findings to the presence of a fraction of small oligomeric aggregates in the highly concentrated protein SAXS sample without cucurbit[8]uril. The aggregation numbers were calculated from the aggregate mass M which was determined from the SAXS data according to:[16]

$$M = I(0)\, N_A / c \Delta\rho_M{}^2,$$

where N_A is Avogadro's number, c (g/cm³) is the protein concentration and $\Delta\rho_M$ (cm/g) is the scattering length difference per mass between solvent and protein. $\Delta\rho_M$ can be calculated from the chemical composition of the protein and the solvent and the specific volume of the protein in solution, which was assumed to be 0.7425 cm³/g, a typical value for globular proteins.[17]

Table 1 Radius of gyration R_G, forward scattering intensities $I(0)$ and aggregation numbers N.

Sample	R_G [nm]	$I(0)$ [cm^{-1}]	N
Theoretical 1YFP (pdb)	1.8	-	-
FGG-mYFP	2.2	0.6899	1.3
FGG-mYFP + cucurbit[8]uril	2.5	1.0197	2.1

To facilitate the determination of the 3-dimensional structure of the cucurbit[8]uril-induced FGG-mYFP dimer in solution with the software BUNCH,[18] efficient use should be made of the available structural information for the system. Most importantly, this is the structure of YFP as determined by X-ray crystallography (PDB entry 1YFP).[15] The complexation of the flexible N-terminus with cucurbit[8]uril does not change the structure of the beta-barrel subunits dramatically, as the scattering patterns of FGG-mYFP in the presence and absence of curcubit[8]uril overlap in the intermediate and high q-regime (Fig. 2a). These units are thus treated as rigid bodies. Moreover, the FGG motifs of the two proteins were imposed to be in close proximity as these are linked within one cucurbit[8]uril molecule. Thus, the dimer was modelled as two rigid bodies (1YFP) linked together by cucurbit[8]uril in the middle of a 21 amino acid flexible tail represented as a chain of 21 dummy residues.

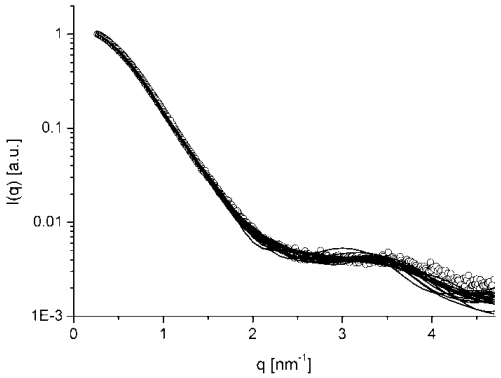

Fig. 3 18 fits with the program BUNCH (lines) to the experimental scattering curve of cucurbit[8]uril-induced FGG-mYFP dimer (open circles).

The program BUNCH[18] was employed to optimize the relative orientations of the rigid bodies and the folding of the tails using a simulated annealing algorithm to give the best agreement between the computed scattering pattern and the experimental SAXS data (Fig. 3). BUNCH performed 18 successive runs on the same data set to generate the averaged structure of the dimer depicted in Figure 4. There is a small, but non-negligible

variation among the resulting structural models which is indicative of a certain degree of flexibility of the system. All of the models correspond to a compact z-shaped structure with a contact area between the YFP rigid bodies. In addition, the long axes of the two YFP barrels are almost parallel to each other which indicates that the z-shaped structure is on average aligned in one plane. This parallel orientation explains the very efficient fluorescence resonance energy transfer observed between these proteins.[9] SAXS experiments have revealed a similar protein conformation for covalently linked chimeric fluorescent proteins,[12] supporting the functionality of the supramolecular linker in inducing efficient protein dimerization and orientation control. Importantly, the SAXS results show that the flexible linker, dimerized via the cucurbit[8]uril, does not induce a large interdomain distance, which would lead to two independently dangling protein barrels. Rather, two well-ordered subunits in close proximity are generated, giving support for the effective and non-intrusive nature of the supramolecular protein dimerization elements.

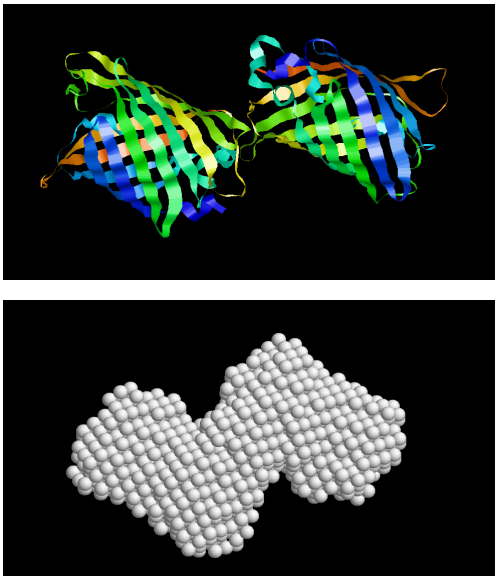

Fig. 4 Results of modelling of the SAXS patterns of FGG-mYFP in the presence of cucurbit[8]uril. Two YFPs are linked together forming a z-shaped structure. top) A typical structure computed as a result of one BUNCH run. bottom) Averaged, most probable structure of a cucurbit[8]uril-induced YFP (FGG-mYFP) dimer computed from 18 runs of the program BUNCH.

Conclusions

Supramolecular, cucurbit[8]uril induced, protein dimerization leads to a well-defined and compact protein dimer, even though the proteins are connected via a flexible and reversible peptide linker. The two supramolecular connected beta-barrels are arranged in a z-shaped structure with the long axes of the two YFP barrels in parallel orientation. This solution structure of a supramolecular protein dimer is the first of its kind. It shows the high functionality of supramolecular systems in controlling protein assembly and aids in the further design and development of functional supramolecular protein architectures.

Notes and references

[a] *Department of Biomedical Engineering, Eindhoven University of Technology, Den Dolech 2, 5612 AZ, Eindhoven, the Netherlands. Fax: (+31)402478367; E-mail: l.brunsveld@tue.nl*
[b] *Research & Development Particle Characterization, Anton Paar GmbH, Anton-Paar-Str. 20, A-8054 Graz, Austria.*
[c] *Institute for Complex Molecular Systems, Eindhoven University of Technology, Den Dolech 2, 5612 AZ, Eindhoven, the Netherlands.*
† Electronic Supplementary Information (ESI) available: Experimental procedures for construction of plasmids, protein expression and purification, dynamic light scattering, and small angle X-ray scattering. See DOI: 10.1039/b000000x/
‡ Funded by ERC grant 204554 – SupraChemBio (LB), NWO VENI Grant 700.10.406 (IKV), and Netherlands Organisation for Scientific Research via Gravity program 024.001.035.

1 S. Burazerovic, J. Gradinaru, J. Pierron, T. R. Ward, *Angew. Chem., Int. Ed.* 2007, **46**, 5510-5514; H. Kitagishi, Y. Kakikura, H. Yamaguchi, K. Oohora, A. Harada, T. Hayashi, *Angew. Chem., Int. Ed.* 2009, **48**, 1271-1274.; Q. Li, C. R. So, A. Fegan, V. Cody, M. Sarikaya, D. A. Vallera, C. R. Wagner, *J. Am. Chem. Soc.* 2010, **132**, 17247-17257; K. Oohora, S. Burazerovic, A. Onoda, Y. M. Wilson, T. R. Ward, T. Hayashi, *Angew. Chem., Int. Ed.* 2012, **51**, 3818-3821.
2 K. L. Heredia, D. Bontempo, T. Ly, J. T. Byers, S. Halstenberg, H. D. Maynard, *J. Am. Chem. Soc.* 2005**, 127**, 16955-16960; C. Boyer, V. Bulmus, J. Q. Liu, T. P. Davis, M. H. Stenzel, C. Barner-Kowollik, *J. Am. Chem. Soc.* 2007**, 129**, 7145-7154; S. Le Gac, E. Schwartz, M. Koepf, J. J. L. M. Cornelissen, A. E. Rowan, R. J. M. Nolte, *Chem.-Eur. J.* 2010**, 16**, 6176-6186; V. M. Hernandez-Rocamora, S. W. A. Reulen, B. de Waal, E. W. Meijer, J. M. Sanz, M. Merkx, *Chem. Commun.* 2011**, 47**, 5997-5999; K. Oohora, A. Onoda, H. Kitagishi, H. Yamaguchi, A. Harada, T. Hayashi, *Chem. Sci.* 2011, **2**, 1033-1038.
3 D. A. Uhlenheuer, K. Petkau, L. Brunsveld, *Chem. Soc. Rev.* 2010**, 39**, 2817-2826.
4 L. Zhang, Y. W. Wu, L. Brunsveld, *Angew. Chem. Int. Ed.* 2007**, 46**, 1798-1802; S. Sakamoto, K. Kudo, *J. Am. Chem. Soc.* 2008**, 130**, 9574-9582; D. A. Uhlenheuer, J. F. Young, H. D. Nguyen, M. Scheepstra, L. Brunsveld, *Chem. Commun.* 2011**, 47**, 6798-6800.
5 J. M. Chinai, A. B. Taylor, L. M. Ryno, N. D. Hargreaves, C. A. Morris, P. J. Hart, A. R. Urbach, *J. Am. Chem. Soc.* 2011, **133**, 8810-8813; R. E. McGovern, H. Fernandes, A. R. Khan, N. P. Power, P. B. Crowley, *Nat. Chem.* 2012, **4**, 527-533; D. Bier, R. Rose, K. Bravo-Rodriguez, M. Bartel, J. M. Ramirez-Anguita, S. Dutt, C. Wilch, F. G. Klärner, E. Sanchez-Garcia, T. Schrader, C. Ottmann, *Nat. Chem.* 2013, **5**, 234-239.
6 P. B. Crowley, P. Ganji, H. Ibrahim, *ChemBioChem* 2008, **9**, 1029-1033; O. Kokhan, N. Ponomarenko, P. R. Pokkuluri, M. Schiffer, D. M. Tiede, *Biochemistry*, 2014, **53**, 5070-5079.
7 C. X. Hou, J. X. Li, L. L. Zhao, W. Zhang, Q. Luo, Z. Y. Dong, J. Y. Xu, J. Q. Liu, *Angew. Chem. Int. Ed.* 2013, **52**, 5590-5593; D. T. Dang, H. D. Nguyen, M. Merkx, L. Brunsveld, *Angew. Chem. Int. Ed.* 2013, **52**, 2915-2919.
8 L. M. Heitmann, A. B. Taylor, P. J. Hart, A. R. Urbach, *J. Am. Chem. Soc.* 2006**, 128**, 12574-12581; J. J. Reczek, A. A. Kennedy, B. T. Halbert, A. R. Urbach, *J. Am. Chem. Soc.* 2009**, 131**, 2408-2415.
9 H. D. Nguyen, D. T. Dang, J. L. J. van Dongen, L. Brunsveld, *Angew. Chem. Int. Ed.* 2010**, 49**, 895-898.
10 D. T. Dang, J. Schill, L. Brunsveld, *Chem. Sci.* 2012**, 3**, 2679-2684.
11 S. Liu, C. Ruspic, P. Mukhopadhyay, S. Chakrabarti, P. Y. Zavalij, L. Isaacs *J. Am. Chem. Soc.* 2005**, 127**, 15959-15967.
12 R. Arai, W. Wriggers, Y. Nishikawa, T. Nagamune, T. Fujisawa, *Proteins* 2004**, 57**, 829-838.
13 J. Lipfert, S. Doniach, *Annu. Rev. Biophys. Biomol. Struct.* 2007**, 36**, 307-327.
14 D. I. Svergun, *Biol. Chem.* 2010**, 391**, 737-743.
15 R. M. Wachter, M. A. Elsliger, K. Kallio, G. T. Hanson, S. J. Remington, *Struct.* 1998**, 6**, 1267-1277.
16 O. Kratky, G. Porod, L. Kahovec, *Z. Elektrochem.* 1951**, 55**, 53-59.
17 E. Mylonas, D. I. Svergun, *J. Appl. Crystallogr.* 2007**, 40**, S245-S249.

[18] M. V. Petoukhov, D. I. Svergun, *Biophys. J.* 2005**, 89**, 1237-1250.