Characteristics of DNA superstructures formed upon interaction with feast/famine regulatory proteins, FFRPs

By Masashi SUZUKI^{†)} and Hideaki KOIKE

National Institute of Advanced Industrial Science and Technology (AIST)

AIST Tsukuba Center 6-10, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566

(Communicated by Masanori OTSUKA, M. J. A., April 14, 2003)

Abstract: By analyzing the 3D structures of full length FFRPs, which were crystallized previously in the absence of DNA, characteristics of DNA superstructures formed upon interaction with these proteins have been predicted. A pair of α -helices 3 in an FFRP dimer are positioned, so that they will bind to two DNA sites separated by 9, or a slightly fewer, basepairs. The nucleotide sequences of the two DNA sites will be reverse complementary to each other, reflecting the symmetric positioning of the two recognition helices. With this DNA fragment embedded in the center of a unit of ~20 basepairs, four such units, altogether of ~80 basepairs, will circle around the disk-like assembly of FFRP. Around each helical turn of the cylindrical assembly of FFRP, six such units, i.e. ~120 basepairs, would wrap.

Key words: AsnC; DNA-binding domain; DNA recognition; FFRP; Lrp; multi-helical; transcription factor.

Introduction. The classification feast/famine regulatory protein (FFRP) groups a number of potential and eu-transcription factors of eubacteria and archaea.¹⁾ Among these the two best characterized are *E. coli* Lrp (leucine-responsive regulatory protein),^{2,3)} and AsnC (asparagine synthesis *C* gene product).⁴⁾ Many aspects of regulations by the two proteins can be understood by the principle of "feast or famine" adaptation for survival of *E. coli* in nutrient enriched or depleted environments.

Many FFRPs have ~160 amino acid residues, i.e. full length (fl) FFRPs. Two such archaeal fl-FFRPs, PflrpA (Pf1493159) from the hyper-thermophilic archaeon *Pyrococcus furiosus*⁵⁾ and pot0434017 from another hyper-thermophilic archaeon *Pyrococcus sp.* OT3,¹⁾ have been crystallized. A fl-FFRP is composed of an N-terminal DNA-binding domain and a C-terminal assembly domain.

Genes of some other FFRPs, archaeal and eubacterial, code only ~80 amino acid residues corresponding to the C-terminal half only of a fl-FFRP (i.e. C-demi-FFRPs).^{1),6)} An example is pot1216151 from *P. OT3*,

which has been crystallized.⁶⁾ In contrast Bs0711124 from the eubacterium *Bacillus subtilis*, and archaeal proteins including Lrs14 (Ss0958260)⁷⁾ from *Sulfolobus solfataricus*, correspond to the N-terminal half of a fl-FFRP, with most of its C-terminus deleted (i.e. N-demi-FFRPs).⁸⁾

The crystal structure of PflrpA (Pf1493159) was determined by another group,⁵⁾ and the structures of pot1216151⁶⁾ and pot0434017¹⁾ by our group. In those of the two fl-FFRPs pot0434017¹⁾ and PflrpA (Pf1493159)⁵⁾ inside their DNA-binding domains (DBDs) sets of three α -helices are combined⁹⁾ in essentially the same way as in many other DNA-binding proteins.⁹⁾⁻¹⁴⁾ In such a multi-helical formation one of the α -helices, usually the third helix, is used for recognizing DNA.⁹⁾⁻¹⁴⁾ It is highly likely that this scheme applies on the design of FFRPs as well,⁸⁾ although no DNA complex of FFRP has been crystallized.

The structural units of the known FFRP assemblies are dimers. By further combining four dimers each of pot1216151⁶⁾ or PflrpA (Pf1493159),⁵⁾ disks are made, while pot0434017 forms cylinders with each helical turn composed of six dimers.¹⁾ The dimer structure is kept essentially the same independent of the assembly forms. By analyzing these crystal structures, in this

^{†)} Correspondence to: M. Suzuki at AIST.

Fig. 1. The ribbon diagram of the CAP dimer in complex with DNA (a and c, PDB code 1CGP), in comparison with that of the FFRP dimer (Pflrpa, Pf1493159) crystallized into the disk-like assembly (b and d, PDB code, 1H1G). Side views (a, b), and top views (c, d) looking down along the helix-axes of the DNA binding (c) or modeled to be binding (d). In each protein dimer one of the two monomers is shadowed. In a the structure of the DNA co-crystallized is shown together with CAP, but from c it is omitted. The central ~20 basepairs of this DNA are overlaid on FFRP in b. Compare Fig. 1b with Fig. 2b for better understanding. The scale 34 Å is shown in a.

paper the essential characteristics of DNA-FFRP complexes are investigated.

The mode of DNA-binding by the FFRP dimer. In Fig. 1 the crystal structure of the FFRP dimer in the disk-form assembly⁵⁾ (Fig. 1b) is compared with that of the CAP (catabolite activator protein) dimer in complex with DNA (Fig. 1a); the structure of the DNA-CAP complex was determined previously by another group.¹⁵⁾ The CAP protein does not belong to the FFRP

group, but its DBD is of the multi-helical type, with the third α -helix being the recognition helix. This protein forms a dimer like fl-FFRPs. Moreover, CAP induces sharp bending of the DNA around this protein, as is expected when DNA wraps around the FFRP assemblies in both cylindrical and disk-like assemblies. The curvature of the CAP-binding DNA is one of the highest among those of the known DNA superstructures, reaching the limit of the deformability of DNA.

Fig. 2. Schematic drawing of pairs of α -helices binding two sites in the DNA major (M) groove, separated by 10 (a) or 9-8 (b) basepairs. Note the phase in the helical periodicity at each site in the DNA major groove.

A pair of recognition helices 3 in the CAP dimer contact two sites in DNA in the major groove (Fig. 1a). The protein dimer faces essentially one side of the DNA, and so the centers of the two DNA sites are positioned essentially in the same phase of the helical periodicity, separated by ~ 10 basepairs^{15,16}; with this number a canonical DNA would span $34 \text{ \AA}$, forming a complete one helical turn.

The two recognition helices in the CAP dimer are nearly parallel to each other (Fig. 1a). The ω angle, which is a measure of the tilt between the two helices when projected along the DNA helix axis, is very small; see Fig. 1c of this paper, and also see Fig. 2c of ref. 16) for the definition of the ω angle.

The separation between the pair of α -helices 3 in the FFRP dimer is smaller than that in the CAP dimer (Fig. 1b, also see Fig. 4 of ref. 8)). In addition, the α -helices 3 are less parallel to each other, creating a larger ω angle (Fig. 1d). All these findings coincide with the idea that when the FFRP dimer binds DNA it contacts two sites separated by 9, or a slightly fewer, basepairs (compare Fig. 2b with Fig. 2a for better understanding).

Each DNA site bound by an α -helix normally spans 4-5 basepairs^{16,17}; a straight α -helix can bind essentially one side of DNA only, and cannot further follow the curved surface of the DNA helix. Thus, DNA bound by an

FFRP dimer will have a sequence such as $[\text{ABCD}]\text{N}_{-5}[\text{DCBA}]$, where $\underline{\text{A}}$, for example, is the base complementary to A, and N is an unspecified nucleotide.

The mode of DNA-binding by the FFRP disk.

By using the crystal structure of the CAP-binding DNA, we now are able to model DNA-FFRP complexes. Here our aim is to understand some essential structural characteristics of DNA-FFRP interactions, for instance, the curvature of the DNA superstructure formed, the approximate number of basepairs wrapping around the protein assemblies, and the diameter of the complexes. Further details or detailed conformations of the DNA sites originated in particular nucleotide sequences, which are further modified upon interaction with FFRP proteins, are not considered.

In Fig. 3a four copies of the 3D structure of the central ~ 20 basepairs in the CAP-binding DNA (Fig. 1a) are overlaid on the four dimers of PflRrpA (Pfl493159)⁵ in the disk-like assembly. Here the binding geometry of each α -helix 3 to DNA is kept the same as in Fig. 1b, which is modeled essentially into the DNA-binding mode shown in Fig. 2b. Being separated by ~ 20 basepairs, the two ends of each DNA fragment are in the same phase of the helical periodicity, and so is the set of the closest ends of each neighboring DNA pair. In order to make a realistic model of the complex each set needs to be connected. In fact, they are not so far away, and not much space is left in between; if basepairs were to be inserted, their numbers should be large, i.e. ~ 10 , ~ 20 , ~ 30 , since the helical periodicity of the DNA needs to be kept, and the Watson strand cannot be connected with the Crick strand.

The four DNA fragments will become close enough, by changing the protein structure rather than by changing the DNA conformation. Namely, the diameter of the protein assembly should decrease by $\sim 10\%$, while keeping the DNA fragments attached to the DBDs (Fig. 4a). This can be achieved by further folding the linker connecting the two domains in each FFRP, although the details of the folded conformation are not specified in this paper (Fig. 4a). This linker seems to be flexible, and not much interaction takes place between the two domains in the two crystal structures of fl-FFRPs.

In short, the DNA that binds around the FFRP disk will be of ~ 80 basepairs, i.e. four repeats of ~ 20 bases in the sequence $\text{N}_{3-4}[\text{ABCD}]\text{N}_{-5}[\text{DCBA}]\text{N}_{3-4}$. The diameter of the disk with the DNA circling around will be $\sim 110 \text{ \AA}$. This value is close to that of nucleosome, where DNA of ~ 160 basepairs creates sharp bending in succession, by forming two left-handed superhelical turns around the

Fig. 3. The crystal structures of the disk-like assembly of pot0434017 (a) and of a full turn in the cylindrical assembly of Pflrpa (Pf1493159) (b), with copies of the ~20 basepair CAP-binding DNA overlaid. In a four copies of the DNA pieces are overlaid on the four FFRP dimers in the same way as in Fig. 1b. In b six copies of the same DNA are positioned, so that these DNA pieces become close to each other, as if they are continuing beyond the gaps.

Fig. 4. Structures of DNA-complexes of the FFRP disk (a) and cylinder (b) modeled. These models are made by assuming that the linker connecting the N-terminal DBD and the C-terminal assembly domain in each FFRP is flexible; in a it is more folded, and in b it is more extended (schematically shown by broken lines). Dimers of FFRP are colored differently.

protein core.¹⁸⁾

The mode of DNA-binding by the FFRP cylinder. Fig. 3b shows the crystal structure of one helical turn composed of six dimers of pot0434017 assembled into the cylinder.¹⁾ On top of this structure six copies of the ~20 basepair CAP-binding site are overlaid. These pieces of DNA are placed, so that their ends become close enough to each other, as if the DNA is continuing crossing over the gaps.

In order to bind this superstructured DNA, the diameter of the protein assembly needs to increase by ~5% (Fig. 4b). It seems reasonable to expect that the flexibility of the protein linker can adjust the geometry of the DBDs, this time by further unfolding and extending (Fig. 4b). Insertion of ~10, ~20 basepairs between the DNA fragments appears to be difficult, since this will further increase the diameter of the DNA superhelix. Deletion of ~10 basepairs from each DNA fragment (i.e. from ~5 basepairs from each end) does not seem to be possible, since this implies that a single site in the DNA needs to be shared by two α -helices approaching from two FFRP dimers positioned nearby.

In short, ~160 basepairs of DNA (i.e. six repeats of the ~20 basepairs) will bind around each helical turn of the cylindrical FFRP assembly; $(N_{3-4}[ABCD]N_{-5}[DCBA]N_{3-4})_6$. The diameter of the DNA-FFRP complex will be ~160 Å.

Conclusions. It is likely that upon binding to DNA the conformation of the linker connecting the N- and C-domains of each FFRP is altered, thereby adjusting the geometry of the set of DBDs in the assembly. The nucleotide sequences recognized by FFRP assemblies, of which the units are dimers, will be repeats of ~20 basepairs, $N_{3-4}[ABCD]N_{-5}[DCBA]N_{3-4}$. Such regularity is not expected for the sites recognized by FFRPs lacking the assembly domain (i.e. N-demi-FFRPs). Interpretation of physico-chemical and biochemical data on DNA-binding by FFRPs (see examples in refs. 19) and 20)) need to be carried out carefully with considering the above findings.

Acknowledgements. This work was supported by the Core Research for Evolutional Science and Technology (CREST) program of the Japan Science

and Technology Corporation (JST), New Energy and Industrial Technology Development Organization (NEDO), and the Mitsubishi Foundation. We thank Dr. Lester Clowney for his critical reading of the manuscripts.

References

- 1) Koike, H., Sakuma, M., Mikami, A., Amano, N., and Suzuki, M. (2003) *Proc. Japan Acad.* **79B**, 65-71.
- 2) Calvo, J. M., and Matthews, R. G. (1994) *Microbiol. Rev.* **58**, 466-490.
- 3) Newman, E. B., and Lin, R. (1995) *Annu. Rev. Microbiol.* **49**, 747-775.
- 4) Kölling, R., and Lother, H. (1985) *J. Bacteriol.* **164**, 310-315.
- 5) Leonard, P. M., Smits, S. H. J., Sedelnikova, S. E., Brinkman, A. B., de Vos, W. M., van der Oost, J., Rice, D. W., and Rafferty, J. B. (2001) *EMBO J.* **20**, 990-997.
- 6) Kudo, N., Allen, M. D., Koike, H., Katsuya, Y., and Suzuki, M. (2001) *Acta Crystallogr.* **D57**, 469-471.
- 7) Napoli, A., van der Oost, J., Sensen, C. W., Charlebois, R. L., Rossi, M., and Claramella, C. W. (1999) *J. Bacteriol.* **181**, 1474-1480.
- 8) Suzuki, M., Amano, N., and Koike, H. (2003) *Proc. Japan Acad.* **79B**, 92-98.
- 9) Suzuki, M., and Brenner, S. E. (1995) *FEBS Lett.* **372**, 215-221.
- 10) Suzuki, M. (1995) *Proc. Japan Acad.* **71B**, 27-31.
- 11) Suzuki, M., and Makino, K. (1995) *Proc. Japan Acad.* **71B**, 132-137.
- 12) Suzuki, M., Masumoto, H., and Okazaki, T. (1995) *Proc. Japan Acad.* **71B**, 153-157.
- 13) Nitta, Y., and Suzuki, M. (1995) *Proc. Japan Acad.* **71B**, 193-197.
- 14) Suzuki, M., Suckow, J., Kisters-Woike, B., Aramaki, H., and Makino, K. (1996) *Adv. Biophys.* **32**, 31-52.
- 15) Schulz, S. C., Shields, G. C., and Steitz, T. A. (1991) *Science* **253**, 1001-1001.
- 16) Suzuki, M., Yagi, N., and Gerstein, M. (1995) *Protein Engineer.* **8**, 329-338.
- 17) Suzuki, M., and Gerstein, M. (1995) *Proteins* **23**, 525-535.
- 18) Luger, K., Mäder, A. W., Richmond, R., Sargent, D. F., and Richmond, T. J. (1997) *Nature* **389**, 251-260.
- 19) Wang, Q., and Calvo, J. M. (1993) *J. Mol. Biol.* **229**, 306-318.
- 20) Wang, Q., and Calvo, J. M. (1993) *EMBO J.* **12**, 2495-2501.