

Structural bioinformatics

Structural drift: a possible path to protein fold change

S. Sri Krishna² and Nick V. Grishin^{1,2,*}¹Howard Hughes Medical Institute and ²Department of Biochemistry, University of Texas Southwestern Medical Center, 5323, Harry Hines Blvd, Dallas, TX 75390-9050, USA

Received on October 21, 2004; revised on November 22, 2004, accepted on December 13, 2004

Advance Access publication December 16, 2004

ABSTRACT

Summary: Along with their mutating sequences, protein structures change in time. Analyzing a formate dehydrogenase domain that is evolutionarily related to ferredoxin, but simultaneously contains all the structural elements of a β -Grasp fold, we illustrate here a mechanism termed as structural drift, by which changes to a protein fold can occur.

Contact: grishin@chop.swmed.edu

It is commonly accepted that spatial structures of proteins are more conserved than their sequences. Many evolutionarily related proteins with pronounced structural resemblance and undetectable sequence similarity are known (Andreeva *et al.*, 2004). Often, structural features remain preserved even after sequence signal is lost to mutations, insertions and deletions. With more protein structures being solved and compared with each other, we are beginning to see many exceptions to this general rule (Andreeva *et al.*, 2004; Grishin, 2001). Moreover we are starting to catalogue possible evolutionary scenarios that lead to significant structural changes between homologues (Grishin, 2001). Here we discuss an example of how such changes can occur.

The modular and repetitive nature of protein structures has been well characterized by Thornton and colleagues as the Russian doll effect, namely when smaller substructures are contained within larger structures (Swindells *et al.*, 1998). In proteins that display the Russian doll effect, there are multiple levels of increased structural complexity. The protein at every new level is formed by the addition of secondary structural elements to the core of the previous level (Swindells *et al.*, 1998). However, the common minimum core (smallest doll) always remains the same by definition, and thus the fold does not change.

As an alternative to this description, one can imagine an evolutionary scenario in which addition of a few elements to a protein structure may partially or completely disrupt the old core and form a distinct subdomain that includes the newly added elements along with parts of the old core, while not the entire old core like in the Russian doll effect. This newly formed subdomain may share structural similarity to other evolutionarily unrelated proteins. We call such a phenomenon structural drift. A protein experiencing structural drift can be described as a hybrid of two overlapping subdomains, which also have overlapping elements from their respective cores. The first subdomain constitutes the old protein core and can share sequence, structural and functional similarities with its homologues. The second subdomain covers several newly added elements and

a part of the old core, and thus overlaps with the first subdomain. This subdomain may show analogous structural resemblance to other folds. Structural drift is a special case of gregariousness, a property described by Harrison *et al.* (2002) that measures how many other folds have a significant structural overlap with a given fold irrespective of evolutionary reasons. We illustrate the structural drift concept by analyzing the ferredoxin domains of formate dehydrogenase (FDH) small subunit (Raaijmakers *et al.*, 2002) and speculate about its evolutionary implications.

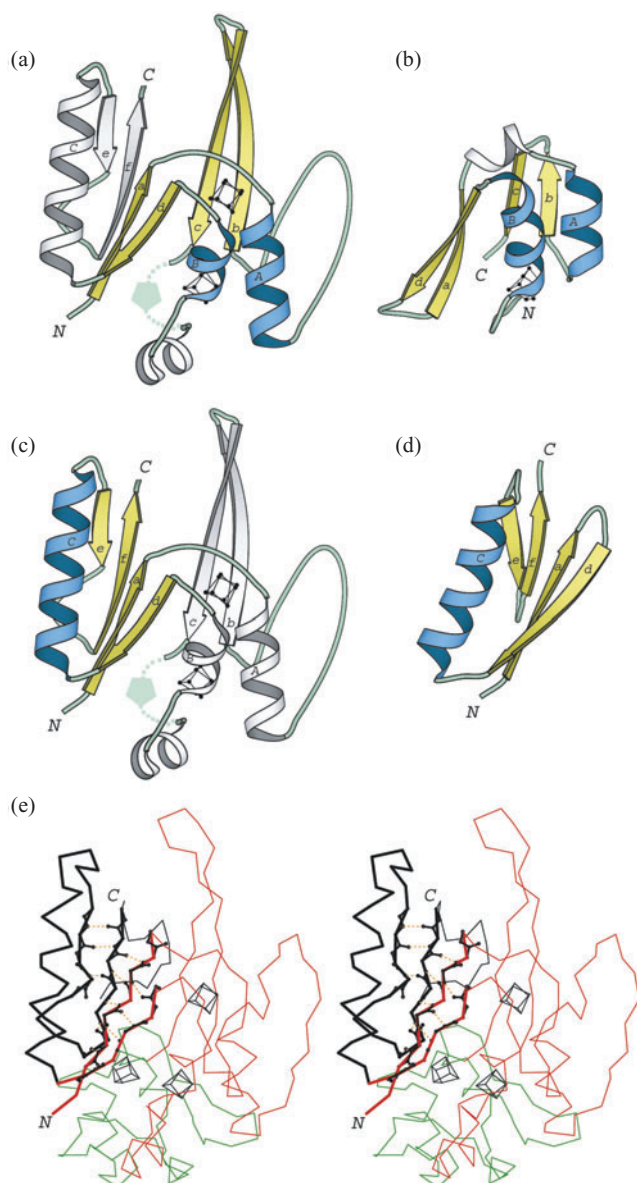
Ferredoxins are small Fe–S proteins functioning as electron carriers in a variety of redox reactions. Ferredoxins are present in all organisms and adopt several different structural folds. The bacterial ferredoxin fold (Andreeva *et al.*, 2004) has a two-layer $\alpha + \beta$ architecture and is made of four β -strands and two α -helices. This α – β plait structure (Orengo *et al.*, 2003) is one of the most common folds in nature. In addition to 4[Fe–S]-ferredoxins, this fold is seen in a functionally diverse set of proteins encompassing 48 superfamilies from the Structural Classification of Proteins (SCOP) database 1.65 (Andreeva *et al.*, 2004).

The 214-residue small subunit of FDH binds three 4[Fe–S] clusters and consists of two homologous ferredoxin domains (Raaijmakers *et al.*, 2002) originated by a duplication event (George *et al.*, 1985) (PDB Entry 1h0h, chain B, Fig. 1a, b and e). In addition to the ferredoxin core structural elements (β -strand *a*, α -helix *A*, β -hairpin *bc*, α -helix *B* and β -strand *d*, Fig. 1a), the first domain contains a C-terminal extension. This extension is comprised of a long α -helix followed by a β -hairpin (elements *C*, *e* and *f*). The second ferredoxin domain (Fig. 1b) is inserted inside the first domain between β -strand *c* and α -helix *B* (Fig. 1a, c and e), circularly permuted with respect to the first domain, and consists of just the ferredoxin core secondary structural elements (*BdaAbc*). Sequence, structural and functional arguments (Raaijmakers *et al.*, 2002) leave little doubt that these two ferredoxin domains of FDH are evolutionarily related (i.e. homologous) and the elements *aAbcBd* form their evolutionarily conserved functional cores that bind 4[Fe–S] clusters. However, the C-terminal extension of the first domain provokes a few thoughts.

This C-terminal extension (Fig. 1a, elements *C*, *e* and *f* shown in white; Fig. 1e traced in black thick lines) together with two β -strands from the ferredoxin core (Fig. 1a, β -strands *a* and *d*; Fig. 1e, red thick lines) forms a substructure (Fig. 1c, colored yellow and blue) that has a pronounced similarity to proteins from the β -Grasp fold. The structural core of β -Grasp (ubiquitin-like) fold is made up of two β -hairpins arranged to form a mixed β -sheet and connected by an α -helix: $\beta_2\alpha\beta_2$ (Fig. 1d). A structure similarity search with the first domain of the FDH small subunit (all elements including the

*To whom correspondence should be addressed.

C-terminal extension, but not the second ferredoxin domain, Fig. 1a and c) against a representative set of SCOP domains using the program MAMMOTH (Ortiz *et al.*, 2002) finds hits to itself ($Z = 20.6$), other ferredoxin domains ($Z = 5.6\text{--}7.8$) and the β -Grasp fold domains ($Z = 5.7$). MAMMOTH Z -scores for the hits to homologous ferredoxins and structurally similar β -Grasp are comparable. The β -Grasp-like core of the FDH small subunit (Fig. 1c, colored yellow and blue) can superimpose with the immunoglobulin light chain-binding domain of protein L (1hz6, Fig. 1d) with an RMSD of 1.27 Å over 52 C_α atoms from all four β -strands (2.34 Å over 84 C_α atoms, including residues from the α -helix). This superposition is comparable to that of the second ferredoxin domain onto the first one, which can be superimposed over 62 C_α atoms with an RMSD of 1.19 Å (1.90 Å over 86 C_α atoms when residues from strands *a* and *d* are included). As demonstrated by structural searches and superpositions, the first domain of the FDH small subunit is gregarious (Harrison *et al.*, 2002) and shows almost equal structural similarity



to the ferredoxin fold (*aAbcBd*) and to the β -Grasp fold (*adCef*). The cores of both folds overlap within the FDH structure, namely, β -strands *a* and *d* belong to both cores simultaneously. Since clear homology exists between the ferredoxin cores (Raaijmakers *et al.*, 2002), the structural similarity to β -Grasp fold is analogous and probably fortuitous due to the overlap with the ferredoxin core over the β -hairpin *ad*, i.e. *aAbcBd* is the evolutionary unit with *Cef* being an extension that disrupts the *aAbcBd* ferredoxin unit by recruiting the *ad* β -hairpin to form a β -Grasp core *adCef*.

It is also noteworthy that the domain delineation program PUU (Holm and Sander, 1994) defines the first domain of the FDH small subunit to cover the elements of a β -Grasp fold rather than a ferredoxin fold. This definition is mostly due to the well-defined hydrogen bonds between the β -strands *a*, *d*, *e* and *f* that constitute the β -sheet in β -Grasp (Fig. 1e, β -strands shown as thick lines and ball-and-stick representation with hydrogen bonds highlighted as orange dashed lines). In contrast, the β -sheet of the ferredoxin fold is irregular (Fig. 1a, colored yellow and blue) and β -strands *a* and *c* are not hydrogen-bonded in both ferredoxin domains (Fig. 1a and b). Such absence of hydrogen bonds is characteristic of the entire superfamily of 4[Fe-S] ferredoxins, although in other superfamilies of the ferredoxin-like fold (e.g. ribosomal protein S6) β -strands *a* and *c* are hydrogen-bonded to form a regular four-stranded β -sheet. The second ferredoxin domain of the FDH small subunit is a typical representative of the 4[Fe-S]-ferredoxin superfamily. It lacks the hydrogen bonds between the β -strands *a* and *c*, but all four β -strands (*a*, *b*, *c* and *d*) are positioned roughly in the same plane (150°) (Fig. 1b). The mutual orientation of the β -hairpins *ad* and *bc* is different in the first FDH small subunit domain. The presence of the C-terminal β -hairpin *ef* twists the *ad* β -hairpin by $\sim 45^\circ$ and causes it to move further away (~ 3 Å) from the *bc* β -hairpin (Fig. 1a) as compared to the second ferredoxin domain (Fig. 1b), and the two

Fig. 1. Structural drift between ferredoxin and β -Grasp folds. (a) The first ferredoxin domain of the small subunit of FDH (PDB Entry 1h0h, chain B). β -Strands shown in yellow and α -helices in blue comprise the core of the ferredoxin fold. An inserted homologous domain (b) that is related to this domain by a circular permutation is shown as a green pentagon. The 4[Fe-S] clusters are shown as ball-and-stick (black). The C-terminal extension to the ferredoxin fold is colored white. In all further panels of this figure, the elements are labelled according to the first ferredoxin domain for clarity. (b) The second ferredoxin domain that is found inserted inside the first domain (a) is shown in an orientation similar to the first domain. (c) The first ferredoxin domain of the small subunit of FDH. The elements that comprise the core of a β -Grasp-like fold are colored; other elements are shown in white. (d) Immunoglobulin-binding domain of protein L (PDB Entry 1hz6) that has a pronounced structural similarity to the β -Grasp-like fold formed by the elements of the first ferredoxin domain and the C-terminal extension. (e) Stereo diagram showing the complete polypeptide chain of the small subunit of FDH. The first ferredoxin domain is colored red, the second (inserted) ferredoxin domain is colored green and the C-terminal extension is colored black. The elements that comprise the core of the β -Grasp-like fold are shown in thick lines. The backbone atoms of the residues from the β -strands are shown in black and the hydrogen bonds between the β -strands are shown as orange lines. All figures were made using the program BOBSCRIPT (Esnouf, 1999).

β -hairpins of the ferredoxin core (*ad* and *bc*) are now positioned at an angle (105°) rather than being in the same plane.

Thus, the presence of the C-terminal extension structurally disrupts the evolutionary ferredoxin core of the first domain making it less regular than that of the second ferredoxin domain. Such a disruption of the ferredoxin core is compensated by the formation of a regular β -Grasp core (Fig. 1c and e). We call this change a structural drift from a classic ferredoxin domain to a β -Grasp domain by insertion (*Cef*), incorporation of the insertion into the core, and formation of a new core resulting in the distortion of the former core. One could imagine a further evolutionary development, in which other elements of the ferredoxin fold (*bc* β -hairpin, α -helices *A* and *B*) may deteriorate after functional diversification of the domain. The resulting domain will then display only the elements of the β -Grasp core. Such a domain will be structurally analogous and evolutionarily unrelated to the β -Grasp fold proteins, since it contains parts of its relative, a ferredoxin domain.

ACKNOWLEDGEMENTS

We are grateful to Lisa Kinch and Sara Cheek for the helpful comments on our manuscript. This work was supported by NIH grant GM67165 to N.V.G.

REFERENCES

- Andreeva,A., Howorth,D., Brenner,S.E., Hubbard,T.J., Chothia,C. and Murzin,A.G. (2004) SCOP database in 2004: refinements integrate structure and sequence family data. *Nucleic Acids Res.*, **32** (Database issue), D226–D229.
- Esnouf,R.M. (1999) Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallogr. D Biol. Crystallogr.*, **55**, 938–940.
- George,D.G., Hunt,L.T., Yeh,L.S. and Barker,W.C. (1985) New perspectives on bacterial ferredoxin evolution. *J. Mol. Evol.*, **22**, 20–31.
- Grishin,N.V. (2001) Fold change in evolution of protein structures. *J. Struct. Biol.*, **134**, 167–185.
- Harrison,A., Pearl,F., Mott,R., Thornton,J. and Orengo,C. (2002) Quantifying the similarities within fold space. *J. Mol. Biol.*, **323**, 909–926.
- Holm,L. and Sander,C. (1994) Parser for protein folding units. *Proteins*, **19**, 256–268.
- Orengo,C.A., Pearl,F.M. and Thornton,J.M. (2003) The CATH domain structure database. *Methods Biochem. Anal.*, **44**, 249–271.
- Ortiz,A.R., Strauss,C.E. and Olmea,O. (2002) MAMMOTH (matching molecular models obtained from theory): an automated method for model comparison. *Protein Sci.*, **11**, 2606–2621.
- Raaijmakers,H., Macieira,S., Dias,J.M., Teixeira,S., Bursakov,S., Huber,R., Moura,J.J., Moura,I. and Romao,M.J. (2002) Gene sequence and the 1.8 Å crystal structure of the tungsten-containing formate dehydrogenase from *Desulfovibrio gigas*. *Structure (Camb.)*, **10**, 1261–1272.
- Swindells,M.B., Orengo,C.A., Jones,D.T., Hutchinson,E.G. and Thornton,J.M. (1998) Contemporary approaches to protein structure classification. *Bioessays*, **20**, 884–891.