SEGMENT SWAPPING AIDED THE EVOLUTION OF ENZYME FUNCTION: THE CASE OF UROPORPHYRINOGEN III SYNTHASE

András Szilágyi[1*], Dániel Györffy[1], Péter Závodszky[1]

[1] Institute of Enzymology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary

*Corresponding author: András Szilágyi, Institute of Enzymology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Magyar tudosok krt. 2., Budapest 1117, Hungary. Email: szilagyi.andras@ttk.mta.hu, Phone: +36 1 382 6717

Short title: Segment swapping and enzyme evolution

Keywords: conformational entropy; protein dynamics; ligand binding; free energy; protein evolution

This is an edited author's version of the final submitted manuscript. This article has been published in Proteins: Structure, Function, and Bioinformatics. See publisher's version at http://dx.doi.org/10.1002/prot.25190

ABSTRACT

In an earlier study, we showed that two-domain segment-swapped proteins can evolve by domain swapping and fusion, resulting in a protein with two linkers connecting its domains. We proposed that a potential evolutionary advantage of this topology may be the restriction of interdomain motions, which may facilitate domain closure by a hinge-like movement, crucial for the function of many enzymes. Here, we test this hypothesis computationally on uroporphyrinogen III synthase, a two-domain segment-swapped enzyme essential in porphyrin metabolism. To compare the interdomain flexibility between the wild-type, segment-swapped enzyme (having two interdomain linkers) and circular permutants of the same enzyme having only one interdomain linker, we performed geometric and molecular dynamics simulations for these species in their ligand-free and ligand-bound forms. We find that in the ligand-free form, interdomain motions in the wild-type enzyme are significantly more restricted than they would be with only one interdomain linker, while the flexibility difference is negligible in the ligand-bound form. We also estimated the entropy costs of ligand binding associated with the interdomain motions, and find that the change in domain connectivity due to segment swapping results in a reduction of this entropy cost, corresponding to ~20% of the total ligand binding free energy. In addition, the restriction of interdomain motions may also help the functional domain-closure motion required for catalysis. This suggests that the evolution of the segment-swapped topology facilitated the evolution of enzyme function for this protein by influencing its dynamic properties.

INTRODUCTION

The appearance of multidomain proteins in protein evolution greatly facilitated the evolution of a wide range of new protein functions and a complex interaction network. Multidomain proteins evolved by duplication, divergence, and recombination of existing protein domains,[1] which led to a diversification of interactions and functions[2, 3]. Segment-swapped proteins[4] are a group of multidomain proteins with a remarkable evolutionary history: they likely evolved from domain-swapped homodimers by the fusion of the subunits and subsequent divergence of the sequences of the domains. Contrary to most other multidomain proteins, segment-swapped proteins have a discontinuous domain, formed from the N- and C-terminal segments of the chain, with another domain inserted between them. Thus, the two domains are connected by two linkers (see "Swapped" proteins in Fig. 1). In an earlier study[4], we found that many segment-swapped proteins bind a ligand in a cleft between the two domains, and the ligand binding involves a hinge-type relative motion of the domains.

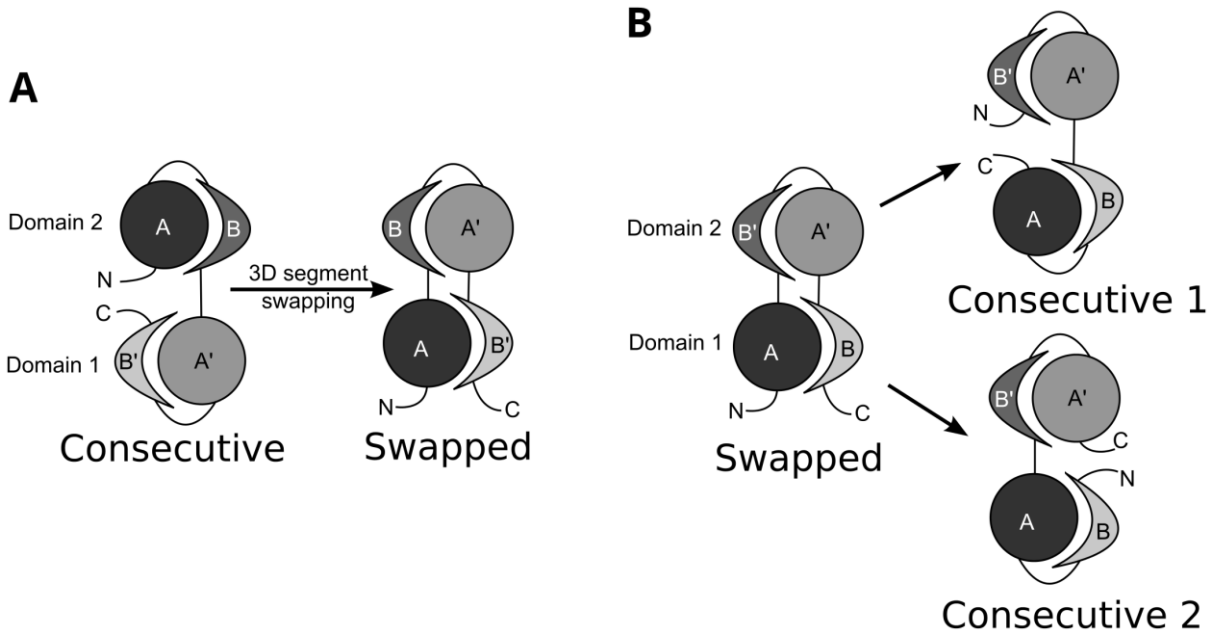

Figure 1 (A) An evolutionary transition of a protein with two consecutive domains into a segment-swapped protein. **(B)** A segment-swapped protein can be converted into a protein with two consecutive domains by circular permutations in two ways, thereby recreating a possible evolutionary ancestor of the segment-swapped protein.

Ligand binding in the cleft between the domains is usually associated with a closure of the domains which fixes their relative orientation, incurring an entropy cost. In our earlier study, we hypothesized that the presence of two linkers between the domains, as opposed to only one linker, restricts relative domain motions in the ligand-free state in a way that may reduce the entropy cost of ligand binding[4]. Thus, the evolutionary process of segment swapping, which transforms consecutively connected domains (with one linker between them) into domains connected with two linkers, may facilitate the evolution of enzyme function, as enzymes are the most prominent examples of proteins binding and acting on ligands.

Uroporphyrinogen III synthase (U3S) is a prime example of a segment-swapped enzyme acting on a ligand that binds between its two domains[5] (Fig. 2). U3S is an essential enzyme present in all three domains of life[6]. It is involved in porphyrin metabolism, including the

biosynthesis of the heme group; it converts hydroxymethylbilane to uroporphyrinogen III. Reduced activity of U3S causes congenital erythropoietic porphyria (CEP), a rare but severe disease. A number of mutations linked with CEP have been identified, by far the most common being the C73R mutation[7]. Recently, it has been shown that position 73 is coupled to the hinge region separating the domains, and mutations at this position modulate the inter-domain closure and affect protein stability[8]. Thus, the hinge region and relative domain movements are critical for the function of U3S.

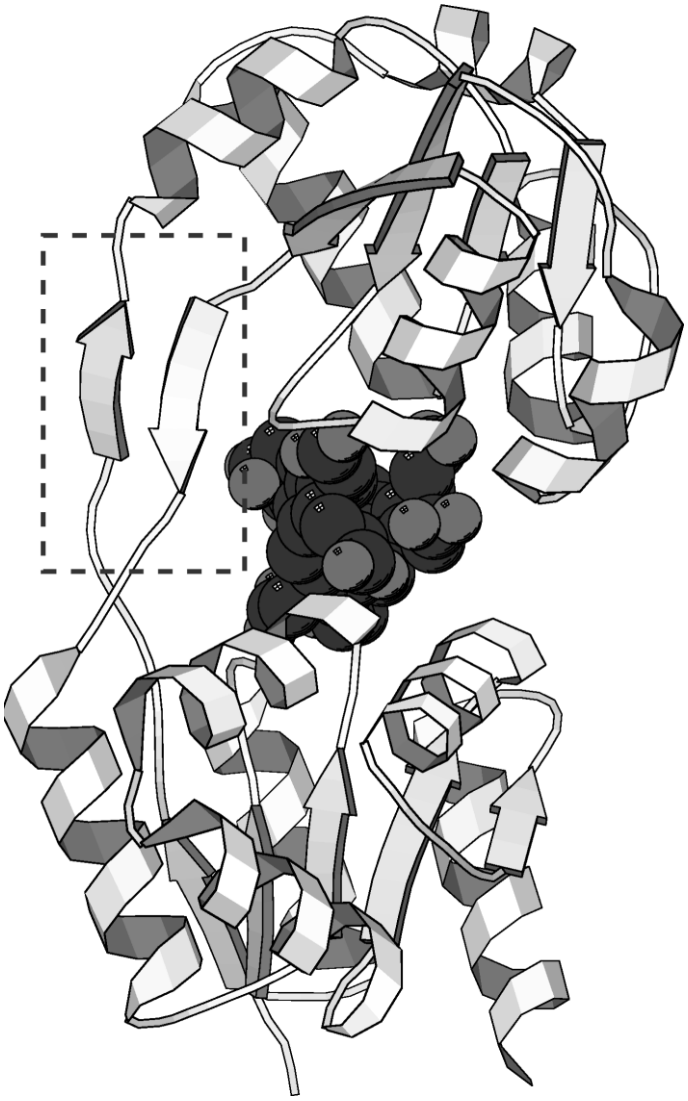

Figure 2 The structure of U3S from *Th. thermophilus* with bound ligand (PDB: 3D8N[5]). The dashed box highlights the linker region that connects the two domains. It consists of a pair of antiparallel β strands; this β sheet is irregular to varying degrees in other structures of U3S. In the circular permutants we constructed for our computational study, 3 residues are deleted from one of the β strands.

U3S consists of two domains linked with two linkers. The structures of both domains are based on a HemD-like fold, a three-layer αβα-sandwich (Fig. 2), but they are circular permutants of each other (see "Swapped" in Fig. 1A). Several lines of evidence suggest that the segment-swapped structure of U3S has evolved by the "domain swapping and fusion" mechanism[4]. The discontinuous domain (i.e. the one comprising the N- and C-termini) has more homologs and structural analogs than the continuous domain, indicating that it is more

ancient[4]. Also, there is a protein, phosphoribosyl pyrophosphate synthetase 1[9], having a similar overall structure to U3S but differing from it in the location where the continuous domain is inserted into the discontinuous domain, indicating a structural variability consistent with the "domain swapping and fusion" mechanism[4]. There are also a number of proteins having two domains with the HemD fold in a consecutive configuration (e.g. diguanylate cyclase[10]).

These observations suggest that the evolutionary history of U3S likely included a stage when its two domains were consecutively connected[4]. This hypothetical ancestor of U3S would have had essentially the same structure as present-day U3S, the only difference being that its domains would have been connected by only a single linker rather than two. This ancestral protein may even have had a similar function to U3S, and could have been able to bind a ligand between its domains. Presumably, the sequences of the two domains in this ancestral "consecutive" protein were still very similar to each other (as they were the result of gene duplication) so that the domains were able to open up and exchange contacts with each other, thus forming a segment-swapped topology (Fig. 1A). There may have been an evolutionary period when the consecutive and segment-swapped topologies coexisted, potentially competing with each other; then, as the sequences of the domains diverged, the segment-swapped variant became stabilized and dominant, completing the evolutionary transition (Fig. 1A).

Clearly, the two forms (consecutive and segment-swapped) of the protein were very similar to each other, the only notable difference being that the consecutive form only had one interdomain linker while the segment-swapped form had two. Why did the segment-swapped form become dominant? Our hypothesis is that the segment-swapped form has an evolutionary advantage over the consecutive form because the two interdomain linkers reduce the interdomain flexibility, which results in a reduced entropy cost of ligand binding, and also facilitates the formation of an efficient hinge mechanism that is essential for the enzyme function. To test this hypothesis computationally, we constructed consecutive versions of U3S by circular permutation *in silico* (as illustrated in Fig. 1B), and performed simulations to compare the interdomain flexibilities between the segment-swapped and consecutive variants. This also allows us to estimate the contribution of the segment-swapped topology to the ligand binding entropy and free energy, and give us an indication whether the change from a consecutive to a segment-swapped topology indeed facilitates the evolution of enzyme function.

METHODS

Protein Data Bank entries 1WCW, 3D8R, and 1JR2 were used as ligand-free U3S structures, and 3D8N was used as a ligand-bound structure. Consecutive variants were constructed by circularly permuting each structure; a 3-residue segment was removed from one of the domain linkers, and the original N- and C-termini were connected by copying the corresponding segment from the other domain. When choosing the 3 residues to be deleted in the linker, care was taken to avoid removing residues involved in ligand binding.

Geometric simulations were performed using the program FRODAN[11], a successor of the earlier FRODA algorithm[12]. This program decomposes a protein structure into rigid clusters using the graph-based FIRST approach[13], and uses momentum perturbation to generate an ensemble of conformations for the molecule while applying constraints for hydrogen bonds and hydrophobic contacts. We used momentum perturbation with fixed constraints in

FRODAN to generate 1,000,000 conformations and saved every 100th frame to obtain 10,000 conformations for each structure.

Molecular dynamics simulations were performed with GROMACS 5.0.2[14] using the CHARMM27 force field. After energy minimization, structures were simulated for 60 ns at 300 K with LINCS bond length constraints[15]. Ligand-free molecules were simulated in GBSA implicit solvent[16] with a 1 nm Coulomb cutoff using a time step of 6 fs with virtual sites and heavy hydrogens[17]. For the ligand-bound 3D8N structure, the ligand (uroporphyrinogen-3) was parameterized with the SwissParam tool[18], and the protein-ligand complex was simulated in explicit TIP3P water with particle mesh Ewald electrostatics [19] with a time step of 2 fs. Ensembles of 20,000 (30,000) frames from the implicit (explicit) solvent MD trajectories were used for entropy calculations.

Relative orientation between domains was represented as Tait—Bryan angles, a version of Euler angles, i.e. 3 angles corresponding to yaw, pitch, and roll. The starting structure of a simulation was considered as reference structure. For each conformation in a trajectory, the first domain was superimposed onto the first domain of the reference structure using least-squares superposition, and the Tait—Bryan angles relative to the reference structure were calculated for the second domain.

Entropies were calculated from the ensembles of Tait—Bryan angles by estimating the trivariate probability density as a Gaussian mixture function using a greedy expectation-maximization algorithm[20]. In this algorithm, the optimum number of Gaussian components is automatically determined. The algorithm was run until convergence, and the obtained log-likelihood was used to calculate the entropy corresponding to the Shannon formula (see Results).

Solvent-accessible surface areas were calculated by the "measure sasa" command in VMD,[21] using a probe radius of 1.4 Å.

RESULTS

In order to compare interdomain flexibilities between the segment-swapped form (having two interdomain linkers) and the consecutive form (having one interdomain linker) of U3S, we performed geometric and molecular dynamics simulations on both the ligand-free and the ligand-bound form of both variants. The consecutive variants were manually constructed *in silico* by rearranging the wild-type structure as indicated in Fig. 1B.

INPUT STRUCTURES

Several known ligand-free structures are available for U3S (PDB entries 1WCW, 3D8R, 3D8S, 3D8T for *Thermus thermophilus* U3S[5], and 1JR2 for human U3S[22]), indicating significant flexibility in relative domain orientations in the ligand-free state[5]. There is only a single structure co-crystallized with bound ligand (PDB entry 3D8N, for *Th. thermophilus* U3S[5]). Although this complex contains the product of the enzyme rather than its substrate, biochemical data and modeling suggest that domain closure occurs upon substrate binding and the enzyme-substrate complex essentially has the same closed conformation as the enzyme-product complex[5]. The A and B rings of the substrate most probably bind to the enzyme in the same manner as in the product, and these rings bind to both domains, producing the closed conformation of the enzyme[5]. Also, the product readily binds to the enzyme and acts as a competitive inhibitor with a similar inhibitory constant (K_i) as the

Michaelis constant (K_m) of the substrate[23]. Thus, the product can be viewed as a substrate analog, and the relative orientation and motion of the domains are expected to be essentially the same in the enzyme-product complex as in the enzyme-substrate complex. The 3D8N structure is therefore appropriate as a ligand-bound structure for our calculations. The ligand-free structure 1WCW is in a rather closed conformation, which is very similar to the ligand-bound 3D8N state, while the other structures (3D8R, 3D8S, and 3D8T) represent more open conformations. The two interdomain linkers tend to be hydrogen-bonded to each other as two β strands, but they are distorted to various extents in the available structures.

SIMULATIONS

Ideally, simulations starting from different ligand-free structures of the same molecule should give identical results. This is, however, not the case, as the different structures represent different sub-states of the native state, and simulations do not ensure perfect sampling of the entire native state. Therefore, we performed simulations for each available structure (both the original segment-swapped structure and its consecutive counterparts constructed according to Fig. 1B), and compared the interdomain flexibilities of the segment-swapped and consecutive variants for each structure. Both geometric[24] and molecular dynamics (MD) simulations were performed; these techniques are complementary because geometric simulation provides a better sampling of the conformational space but does not use a proper force field while MD uses a more accurate force field but samples configurational space less efficiently.

We found considerable relative domain motion in all simulations. Even during the simulations starting from the ligand-free structure 1WCW, which represents a rather closed conformation probably unable to bind a ligand, the domains opened up sufficiently to bind a ligand: the angle between the domains increased by as much as 50 degrees, and the distance between the Cα atoms of Gly111 and Val217 (used here to describe the width of the ligand binding pocket) increased from 7.2 to a maximum of 18.6 Å during the geometric simulations.

COMPARISON OF INTERDOMAIN FLEXIBILITIES

The relative orientation of the two domains was described by three Euler angles corresponding to "yaw", "pitch", and "roll" in an analogy to describing airplane orientations. The conformational ensembles obtained from the simulations were represented in the space of these angles (denoted by φ_1, φ_2, and φ_3). Fig. 3 shows the conformational ensembles obtained from geometric simulations for the ligand-free structure 1WCW in its segment-swapped and consecutive form, in the three-dimensional space of Euler angles (two-dimensional projections are shown for clarity). Although the segment-swapped form itself shows considerable interdomain flexibility (the Euler angles vary from about −50° to 50°), the consecutive form has an even larger flexibility, with the Euler angles varying from about −120° to 120°. Results from other simulations and other ligand-free forms show a similar pattern. In order to quantitatively describe the flexibilities, we estimated conformational entropies from the ensembles.

ESTIMATION OF ENTROPIES OF INTERDOMAIN MOTION

For the entropy calculation, we first estimated the probability densities of the ensembles obtained from the simulations in the space of the three Euler angles as described above. The ensembles were centered around the origin, and trivariate Gaussian mixture functions of the form

$$p(\boldsymbol{\varphi}) = \sum_{i=1}^{k} w_i N(\boldsymbol{\mu}_i, \boldsymbol{\sigma}_i)$$

were fitted onto the ensembles; the vector $\boldsymbol{\varphi}$ represents the three Euler angles φ_1, φ_2, and φ_3, w_i is the weight of i-th component, and $N(\boldsymbol{\mu}_i, \boldsymbol{\sigma}_i)$ is the trivariate normal distribution with mean $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\sigma}_i$. Once the probability density is obtained, the molar entropy can be calculated using the Shannon formula

$$S = R \int p(\boldsymbol{\varphi}) \ln p(\boldsymbol{\varphi})\, d\boldsymbol{\varphi}$$

where R is the universal gas constant and the integration goes from −180° to 180° for all three Euler angles. It should be noted that the entropy values obtained by this formula are unit dependent; entropy differences, however, are unit independent and are therefore physically meaningful.

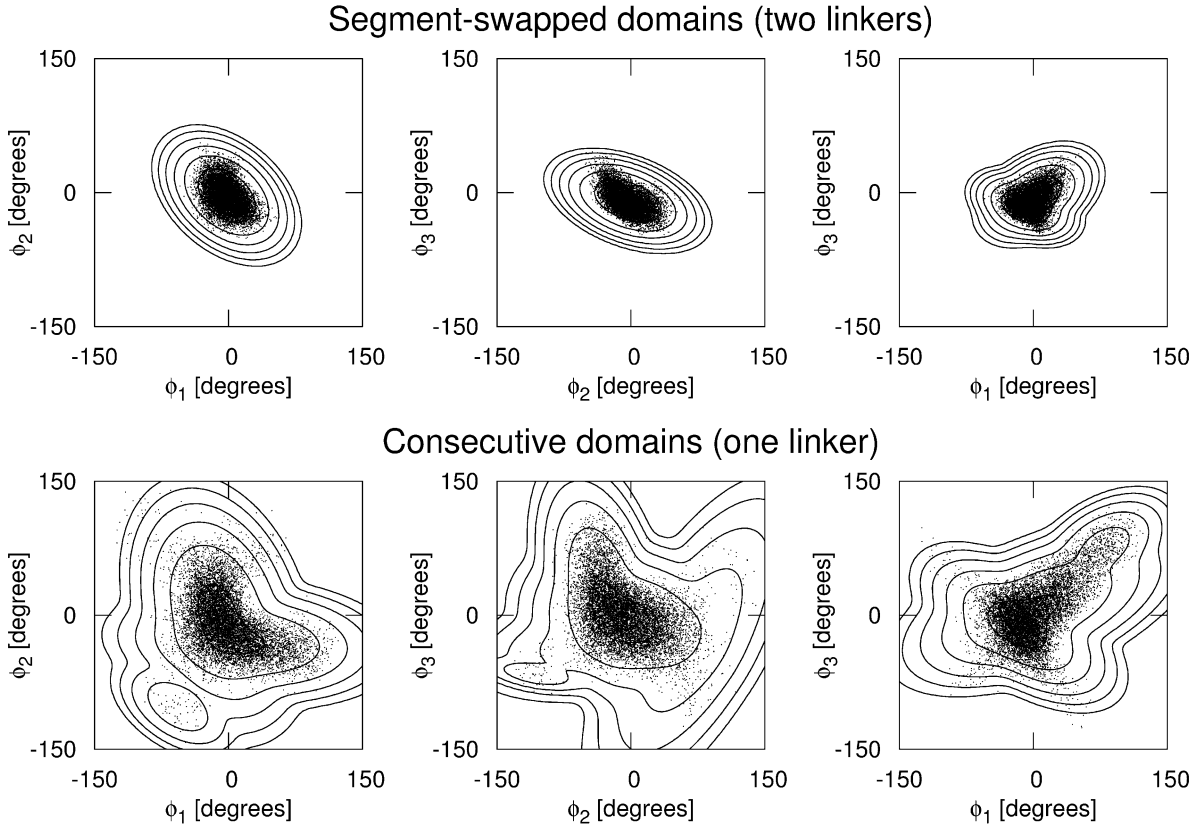

Figure 3 The distribution of the relative domain orientations, described in the space of three Euler angles φ_1, φ_2, and φ_3, obtained from geometric simulations of a segment-swapped U3S (PDB: 1WCW) having two linkers between the domains *(top row)* and a circular permutant with consecutive domains connected by one linker *(bottom row)*. The three two-dimensional projections of the three-dimensional data set are shown for each protein. The dots represent the domain orientations while the contours represent the estimated probability densities obtained by Gaussian mixture fitting.

The plots in Fig. 3 show the fitted probability densities as contour lines. Table 1 shows the calculated interdomain motion entropies from all simulations of all structures. (Results for the *Th. thermophilus* ligand-free structures 3D8R, 3D8S, and 3D8T were almost identical, so we

omitted the results for the latter two.) The entropies for the ligand-free human U3S (1JR2) are also very similar to those for *Th. thermophilus* U3S. Entropies for the ligand-free structure 1WCW were significantly lower than for the other ligand-free structures. We found that the entropies from the MD ensembles tended to be lower than those from the geometric simulation ensembles (indicating narrower sampling by MD), but the differences between the segment-swapped and consecutive variants are consistent and independent of the simulation method.

Table 1 Interdomain motion entropies and entropy differences calculated from simulations of U3S structures. $\Delta S_{swap} = S_{swapped} - S_{consecutive}$ is the domain motion entropy loss due to switching from a consecutive to a segment-swapped topology. $\Delta\Delta S_{bind}$ is the contribution of the segment-swapped topology to the ligand binding entropy as described in the text. Entropies are given in J/K/mol. The numbers are results from geometric simulations except for those in parentheses which come from MD simulations. The data from geometric simulations for the $S_{consecutive}$ column are averages of the two possible consecutive variants.

State	Starting structure (PDB ID)	$S_{consecutive}$	$S_{swapped}$	ΔS_{swap}	$\Delta\Delta S_{bind}$
Bound	3D8N	100.0 (61.6)	98.0 (62.0)	−2.0 (0.4)	-
Unbound	1WCW	116.4 (82.2)	99.1 (64.5)	−17.3 (−17.7)	15.3 (18.1)
	3D8R	134.8 (83.0)	125.4 (71.8)	−9.4 (−11.2)	7.4 (11.6)
	1JR2	134.2	122.2	−12.0	10.0

For the ligand-free structures, the domain-motion associated entropy difference between the consecutive and segment-swapped versions is between 9.4 and 17.3 J/K/mol, indicating that going from a consecutive to a segment-swapped topology significantly reduces relative domain motions.

For the ligand-bound structure 3D8N, the difference between the segment-swapped and consecutive variants is very small, about 2 J/K/mol from the geometric and a negligible 0.4 J/K/mol from the MD simulations, indicating that the bound ligand essentially fixes the relative domain orientation, and cutting one of the linkers between the domains does not have a significant effect in this state.

ESTIMATION OF THE INFLUENCE OF DOMAIN CONNECTIVITY ON LIGAND BINDING ENTROPY

In order to calculate the change in ligand binding entropy due to a switch from consecutive to segment-swapped topology, we rely on the thermodynamic cycle shown in Fig. 4. By taking the differences in the interdomain motion entropy between all four species shown in the scheme (i.e. consecutive and swapped, bound and unbound), we obtain all four entropy changes indicated (i.e. the entropy changes associated with binding in the consecutive and the swapped states, and those associated with segment-swapping in the bound and the unbound states). From these entropy changes, the change in ligand binding entropy due to the switch from consecutive to segment-swapped topology is

$$\Delta\Delta S_{\mathrm{bind}} = \Delta S_{\mathrm{bind,swapped}} - \Delta S_{\mathrm{bind,consec}} = \Delta S_{\mathrm{swap,bound}} - \Delta S_{\mathrm{swap,unbound}}$$

Using the data in Table 1 for the individual species, we find that switching to a segment-swapped topology results in a ligand binding entropy cost reduction of 7.4 to 15.3 J/K/mol (from geometric simulations; MD simulations yield slightly higher values, see Table 1), corresponding to a reduction of 2.22 to 4.59 kJ/mol in ligand binding free energy at 300 K. Although the substrate binding free energy is not known for U3S, Michaelis constants (K_m) for the catalytic reaction range from 1 to 60 μM[8, 25]; if we consider K_m as an approximation to the dissociation constant K_d, this corresponds to a binding free energy of −24 to −34 kJ/mol. *In silico* docking calculations yielded a binding free energy of about −30 kJ/mol for the product and −19 kJ/mol for the inhibitor NMF-bilane[26]. Thus, our results indicate that the segment-swapped topology contributes about 15% to 25% of the ligand binding free energy.

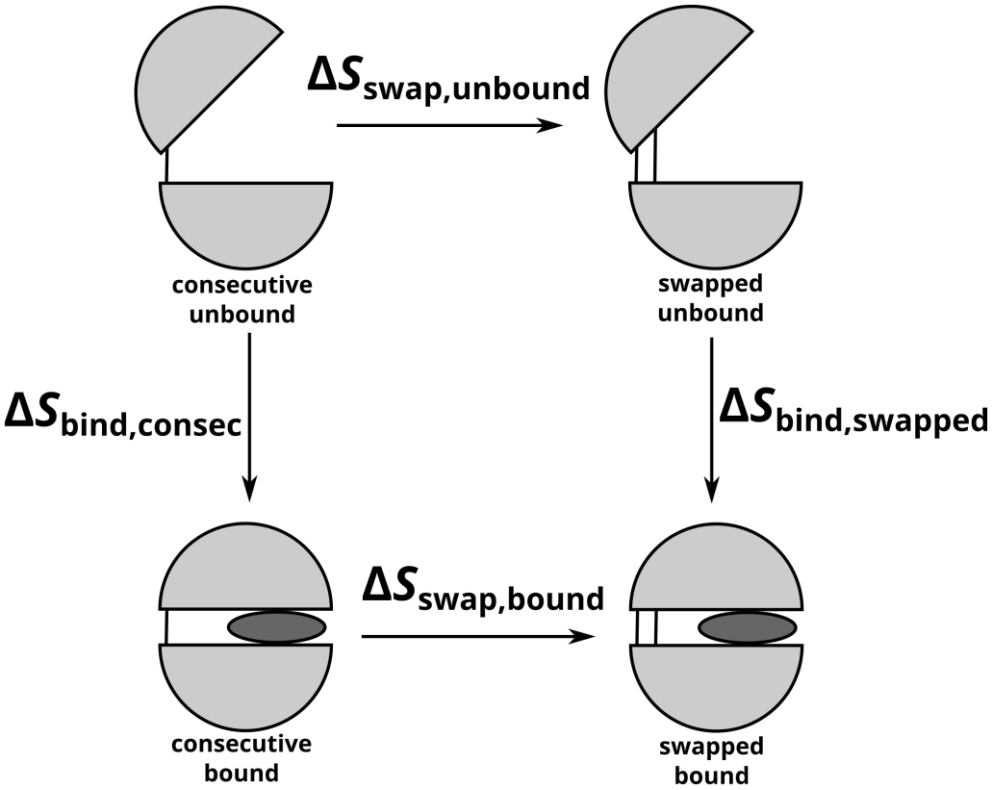

Figure 4 Thermodynamic cycle representing transitions between ligand-bound/unbound and consecutive/swapped states of U3S. The interdomain motion entropy changes are indicated.

OTHER POTENTIAL THERMODYNAMIC EFFECTS OF SEGMENT SWAPPING ON LIGAND BINDING

To check whether segment swapping has any significant impact on the thermodynamics of ligand binding other than through the entropy of interdomain motions, we calculated the solvent-accessible surface area (SASA) for all frames of the trajectories from our simulations. From the geometric simulations, the mean SASA of the molecules was 131/134/145 nm^2 for 3D8N/1WCW/3D8R, and the differences between the segment-swapped and consecutive variants were well within the standard deviation due to fluctuations (~3.3 nm^2). The mean SASA buried upon ligand binding was calculated by taking the differences in SASA between the ligand-free (1WCW, 3D8R) and ligand-bound (3D8N) variants. This was 1.6/3.5 and 13.8/14.6 nm^2 for swapped/consecutive variants of 1WCW and 3D8R, respectively. The

difference between the swapped and consecutive variants in the mean SASA buried upon ligand binding is thus 2.0 and 0.7 nm^2, respectively, which is again well within the standard deviation of the SASA due to fluctuations. These results indicate that segment swapping influences the thermodynamics of ligand binding primarily through restricting the relative motion of the domains while the structure and stability of the domains remain largely intact.

DISCUSSION

The appearance of multidomain proteins in evolution has enabled the emergence of a multitude of new protein functions. Although protein sequence is linear, domains in a multidomain protein are often composed of more than one chain segment, resulting in more than one linker between the domains. We hypothesized that the presence of several linkers restricts relative domain motions, which facilitates the evolution of certain enzyme functions. Here, we tested this hypothesis on U3S. By comparing the original segment-swapped enzyme (having two linkers between the domains) with circularly permuted, consecutive versions (containing only one linker), we have shown that relative domain motions are indeed significantly more restricted in the segment-swapped variant than in the consecutive forms. This reduces the entropy cost and thus the free energy of substrate binding. Although our quantitative results show some variability due to the different starting structures and the different sampling methods, the direction of the change is consistent, and its magnitude is sufficient to provide competitive advantage to the segment-swapped variant. The estimated ligand binding free energy differences could be experimentally verified.

The flexibility restriction due to the presence of two interdomain linkers probably also facilitates the hinge-like, single-axis functional domain-closure motion required for the catalytic action. It should be noted that the restriction of domain motions may also hinder product release, but this can be compensated for by further, simpler evolutionary changes (e.g. point mutations) of the enzyme that reduce the affinity of the product to the protein.

The sequences of the two U3S domains are sufficiently diverged from each other (sequence identities between the domains are 22% for the *Th. thermophilus* enzyme and 13% for the human enzyme) so that the protein cannot “switch back” to a consecutive topology. But at an earlier stage in evolution, the domain sequences must have been much closer to each other, and the consecutive and segment-swapped topologies may have co-existed. In this situation, the segment-swapped version may have had an evolutionary advantage due to the fact that its tighter domain motions made it more suitable to perform an enzyme function.

CONCLUSION

As shown earlier[4], segment swapping is a significant evolutionary route by which multidomain proteins are formed, and many segment-swapped proteins function as enzymes characterized by a hinge-like functional domain motion. We have shown, on the example of U3S, that the change in domain connectivity brought about by segment swapping changes the dynamics of the protein in a way that can facilitate enzyme action. This example illustrates how protein structural topology impacts dynamics and function, and ultimately guides protein evolution.

ACKNOWLEDGMENTS

This work was supported by the Hungarian Scientific Research Fund (OTKA), grant numbers K105415 and NK108642.

REFERENCES

1. Vogel C, Bashton M, Kerrison ND, Chothia C, Teichmann SA. Structure, function and evolution of multidomain proteins. Curr Opin Struct Biol 2004;14(2):208–216.
2. Weiner J, Moore AD, Bornberg-Bauer E. Just how versatile are domains? BMC Evol Biol 2008;8:285.
3. Basu MK, Carmel L, Rogozin IB, Koonin EV. Evolution of protein domain promiscuity in eukaryotes. Genome Res 2008;18(3):449–461.
4. Szilágyi A, Zhang Y, Závodszky P. Intra-chain 3D segment swapping spawns the evolution of new multidomain protein architectures. J Mol Biol 2012;415(1):221–235.
5. Schubert HL, Phillips JD, Heroux A, Hill CP. Structure and mechanistic implications of a uroporphyrinogen III synthase-product complex. Biochemistry 2008;47(33):8648–8655.
6. Storbeck S, Rolfes S, Raux-Deery E, Warren MJ, Jahn D, Layer G. A novel pathway for the biosynthesis of heme in Archaea: genome-based bioinformatic predictions and experimental evidence. Archaea 2010;2010:175050.
7. Fortian A, Castaño D, Ortega G, Laín A, Pons M, Millet O. Uroporphyrinogen III synthase mutations related to congenital erythropoietic porphyria identify a key helix for protein stability. Biochemistry 2009;48(2):454–461.
8. Bdira F ben, González E, Pluta P, Laín A, Sanz-Parra A, Falcon-Perez JM, Millet O. Tuning intracellular homeostasis of human uroporphyrinogen III synthase by enzyme engineering at a single hotspot of congenital erythropoietic porphyria. Hum Mol Genet 2014;23(21):5805–5813.
9. Li S, Lu Y, Peng B, Ding J. Crystal structure of human phosphoribosylpyrophosphate synthetase 1 reveals a novel allosteric site. Biochem J 2007;401(1):39–47.
10. Chan C, Paul R, Samoray D, Amiot NC, Giese B, Jenal U, Schirmer T. Structural basis of activity and allosteric control of diguanylate cyclase. Proc Natl Acad Sci USA 2004;101(49):17084–17089.
11. Farrell DW, Speranskiy K, Thorpe MF. Generating stereochemically acceptable protein pathways. Proteins 2010;78(14):2908–2921.
12. Wells S, Menor S, Hespenheide B, Thorpe MF. Constrained geometric simulation of diffusive motion in proteins. Phys Biol 2005;2(4):S127-136.
13. Jacobs DJ, Rader AJ, Kuhn LA, Thorpe MF. Protein flexibility predictions using graph theory. Proteins 2001;44(2):150–165.
14. Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, Spoel D van der, Hess B, Lindahl E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 2013;29(7):845–854.
15. Hess B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. J Chem Theory Comput 2008;4(1):116–122.
16. Onufriev A, Bashford D, Case DA. Exploring protein native states and large-scale conformational changes with a modified generalized born model. Proteins 2004;55(2):383–394.
17. Feenstra KA, Hess B, Berendsen HJC. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. J Comput Chem 1999;20(8):786–798.
18. Zoete V, Cuendet MA, Grosdidier A, Michielin O. SwissParam: a fast force field generation tool for small organic molecules. J Comput Chem 2011;32(11):2359–2368.
19. Darden T, York D, Pedersen L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J Chem Phys 1993;98(12):10089.
20. Verbeek J, Vlassis N, Kröse B. Efficient greedy learning of Gaussian mixture models. Neural Comput 2003;15(2):469–485.
21. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph 1996;14(1):33–38, 27–28.
22. Mathews MA, Schubert HL, Whitby FG, Alexander KJ, Schadick K, Bergonia HA, Phillips JD, Hill CP. Crystal structure of human uroporphyrinogen III synthase. EMBO J 2001;20(21):5832–5839.
23. Hart GJ, Battersby AR. Purification and properties of uroporphyrinogen III synthase (co-synthetase) from Euglena gracilis. Biochem J 1985;232(1):151–160.
24. Wells SA. Geometric simulation of flexible motion in proteins. Methods Mol Biol 2014;1084:173–192.
25. Omata Y, Sakamoto H, Higashimoto Y, Hayashi S, Noguchi M. Purification and characterization of human uroporphyrinogen III synthase expressed in Escherichia coli. J Biochem 2004;136(2):211–220.
26. Cunha L, Kuti M, Bishop DF, Mezei M, Zeng L, Zhou M-M, Desnick RJ. Human uroporphyrinogen III synthase: NMR-based mapping of the active site. Proteins 2008;71(2):855–873.