Safer gene repair and targeting based on the monomeric meganuclease I-DmoI by design of homologous-recombination-inducing nickase activity

Antonio S. Torralba, Francesco L. Gervasio

Computational Biophysics Group, Structural Biology and Biocomputing Programme
Spanish National Cancer Research Centre (CNIO), C/ Melchor Fdez. Almagro 3, 28029 Madrid, Spain

Homing endonucleases (HEs) are emerging as important tools in gene targeting and repair [1-3]. Desirable properties are their high specificity for long DNA sequences (14-40 base pairs) that are likely to appear very infrequently in whole genomes, and their ability to produce double strand breaks (DSBs) that induce homologous recombination, and therefore can be used to replace a defective gene by a healthy allele. In addition, the modular structure of HEs makes it relatively easy to engineer artificial enzymes that target new, taylored sequences. In fact, members of the LAGLIDADG family of HEs (named after a central structural moiety), in particular the dimeric enzyme I-CreI, have been used to successfully target the gene responsible for the autosomal disease Xeroderma Pigmentosum (XP), both in vitro and in vivo [4,5].

The two main caveats that appear in HEs engineering and use as molecular biology tools are 1) the cytotoxic effects of DSBs, since the break is often repaired by non-homologous end joining (NHEJ), potentially disrupting untargeted genes, and 2) the appearance of undesirable, toxic dimers, when two different monomers need be expressed as part of the modular strategy. Both limitations can be minimized by creating monomeric nickases, i.e. enzymes that produce single strand breaks (SSBs). SSBs are slightly less efficient at inducing homologous recombination, but prevent NHEJ entirely. LAGLIDADG HEs are particularly amenable to the creation of nickases, as they present active sites with pseudo-two-fold symmetry, with two or three metallic centers, so that one pseudo active site can in principle be inactivated selectively. The “nickase strategy” has proved successful in a number of monomeric HEs, including I-SceI [6] and I-AniI [7].

In this work, the most salient structural features of the thermophilic HE I-DmoI [8,9] were studied, in collaboration with Guillermo Montoya’s crystallography group, with a focus on its potential to create a nickase. According to experimental X-ray data, this monomeric enzyme is particular within its HE family in that, in the active form, it only includes two binding sites for Mg2+ (where other enzymes, such as I-CreI, feature three sites), whereas the inactive form with Ca2+ displays only one binding site (where other enzymes have two). A second point of interest was the potential for mutations distant from the active site to induce activity changes at the cleavage site without affecting the binding affinity and sequence recognition. In this respect, allosteric interactions between separate sites in the enzyme were considered.

In this project, computational methods were used. In a first stage, classical molecular dynamics (MD) simulations and quantum mechanical calculations (QM) were used to clarify the role of metallic ions and water at the active site. A typical protocol included the Amber99SB force field, for MD, together with Density Functional Theory (DFT) at the B3LYP/6-31G* level of theory, for QM.

Initial experimental evidence suggested that the traditional central binding site of LAGLIDADG HEs was occupied by a water molecule. However, both MD and QM indicate that this could only occur transiently, as the experimental geometry is incompatible with an energy minimum without a third ion. Indeed, in all calculations with two Mg2+ (and, in fact, Mn2+ and other divalent cations), the distance between the ions consistently decreased by around 2Å from the initial value of slightly over 7Å. Conversely, the experimental structure remained largely stable if the presumed central water was replaced by a third ion, indicating that the vacancies observed in X-ray experiments might be related to low occupancies due to dynamic binding of the ions (Figure 1). To explore this possibility, the binding enthalpies of a series of ions were estimated, based on the relative activity of the enzyme for various cations, which in I-DmoI follows the sequence Mn2+ > Co2+ > Mg2+ >> Ca2+ [10]

According to our results, the cation binding enthalpies when only two ions are present at the active site are ordered (in absolute value) as Co2+ > Mn2+ > Mg2+ > Ca2+. However, when the binding of three ions was considered, the order agrees with the experimental observation, and the calculated sequence was Mn2+ > Co2+ > Mg2+ > Ca2+. Consistently, the inactive Ca2+ has the lowest binding enthalpy in both cases. Furthermore, the binding enthalpy of three Mg2+ ions is closer to that of Ca2+, compared to the equivalent energies for the first two ions, which might explain the

observed experimental structures, were the third ion is lacking. It should be noted that the calculations did not include the entropic contribution, but similar values for all ions are reasonable, so the qualitative conclusions are not expected to change when free energies are considered.

These results indicate, together with the observed structures, that the ion binding could be a limiting step in the activity of the enzyme, and that indeed the central cation is required at some stage of the mechanism, while it might be absent transiently. This conclusion is further supported by the finding of a water file that self-organizes in MD simulations and crosses the active site, providing a mechanism for catalytic water replacement, as well as for ion substitution. From our results, a third ion may play a role in orienting DNA, as it coordinates with the 3G(coding) phosphate [numbering from Ref. 11, which differs by one nucleotide from Ref. 12]. This is in agreement with Ref. 12 and in disagreement with Ref. 11.

In the second part of the project, devoted to identifying target mutations to produce a nickase based on I-DmoI, methods to characterize allosteric interactions within the enzyme were used. Two approaches were attempted. First, Anisotropic Network Models (ANM) were built, from which fluctuation analyses and allosteric couplings were obtained. A new definition of the latter was used [13] which is resilient to small structural changes. The method incorporates standard normal mode information into a procedure to quantify correlations in local distortions in distant regions of a protein, thus producing a measure of allostery. Secondly, allosteric interactions in I-DmoI were explored using Markov State Models (MSM)[14,15]. The method aims at identifying, by means of a suitable clustering, various dynamic states. One advantage of the method is that in principle short simulations can be accumulated to get information of much longer time scales than those present in the simulations.

Allosteric analysis of I-DmoI indicates that the LAGLIDADG helices are anticorrelated to the DNA binding beta-sheets, and that the latter display long range correlations, even between sheets not directly in contact (Figure 2). In particular, beta1 shows strong allosteric interactions with the sheets across the active site, perhaps indicating a mediating role of the catalytic centers in DNA recognition. Our experimental collaborators have indeed tried to selectively inactivate each of the two pseudo active sites, but, even though cleavage preference for one or the other DNA strand was achieved, it was not selective enough to qualify as nickase activity and both strands are ultimately cleaved by all attempted mutants. Furthermore, unlike other monomeric HEs of the same family (I-SceI, I-AniI), it seems impossible to sufficiently decouple DNA binding from DNA hydrolysis, in line with our theorerical results. However, MSM models were not conclusive as to the number of states that I-DmoI populates, and there is still the possibility that some decoupled modes exist. Futher MD simulations would be necessary to clarify this final point.

In summary, in this project the role of a central cation in the I-DmoI active site was explored. It was concluded that the ion is necessary at least transiently for structure stability and most likely activity. Normal mode analysis revealed strong anticorrelations between the DNA binding beta sheets and the active-site-bearing LAGLIDADG moieties, which might explain the experimental difficulties encountered in creating a nickase based in I-DmoI.

Acknowledgements

The Mare Nostrum (Spain), Curie (France) and HECToR (United Kingdom) supercomputer systems, all members of the European PRACE supercomputing network, kindly provided essential computer time. The CNIO HPC cluster was also instrumental during the development of the present project.

[1] M. Hafez and G. Hausner, Genome (2012) 55, 553-569
[2] F. Pâques and P. Duchateau, Curr. Gene Ther. (2007) 7, 49-66
[3] G. Silva et al., Curr. Gene Ther. (2011) 11, 11-27
[4] S. Arnould et al., J. Mol. Biol. (2007) 371, 49-65
[5] P. Redondo et al., Nature (2008) 456, 107-111
[6] Y. Niu, K. Tenney, H. Li, and F. S. Gimble, J. Mol. Biol. (2008) 382, 188-202
[7] A. McConnell Smith et al., Proc. Natl. Acad. Sci. USA (2009) 106, 5099-5104
[8] G.H. Silva, J.Z. Dalgaard, M. Belfort and P. Van Roey, J. Mol. Biol. (1999) 286, 1123-1136
[9] M. J. Marcaida et al., Proc. Natl. Acad. Sci. USA (2008) 105, 16888-16893
[10] C. Aagaard, M. J. Awayez, and R. A. Garrett, Nucleic Acids Res. (1997) 25, 1523-1530
[11] M. J. Marcaida et al., Proc. Natl. Acad. Sci. USA (2008) 105, 16888-16893
[12] C. Aagaard, M. J. Awayez, and R. A. Garrett, Nucleic Acids Res. (1997) 25, 1523-1530
[13] I.A. Balabin, W. Yang and D.N. Beratan, Proc. Natl. Acad. Sci. USA (2009) 106, 14253-14258
[14] G.R. Bowman et al. J. Chem. Phys. (2009) 131, 124101(11pp)
[15] G.R. Bowman, X. Huang and V.S. Pande, Methods (2009) 49, 197-201

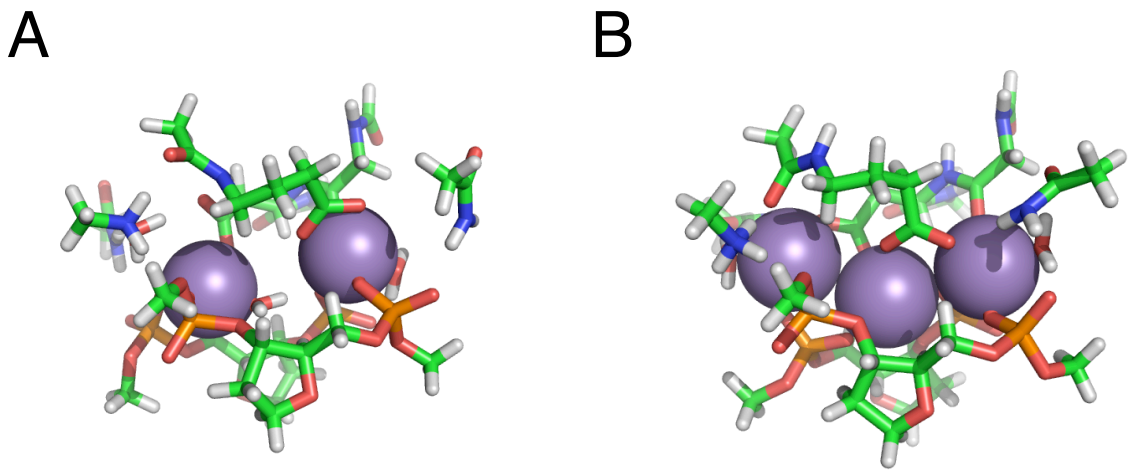

Figure 1. Models of the I-DmoI active site. A. When only two ions (spheres) are considered, as suggested by the experimental structures, the active site contracts and becomes highly asymmetric. B. Conversely, the experimental geometry is better reproduced by a three-ion model

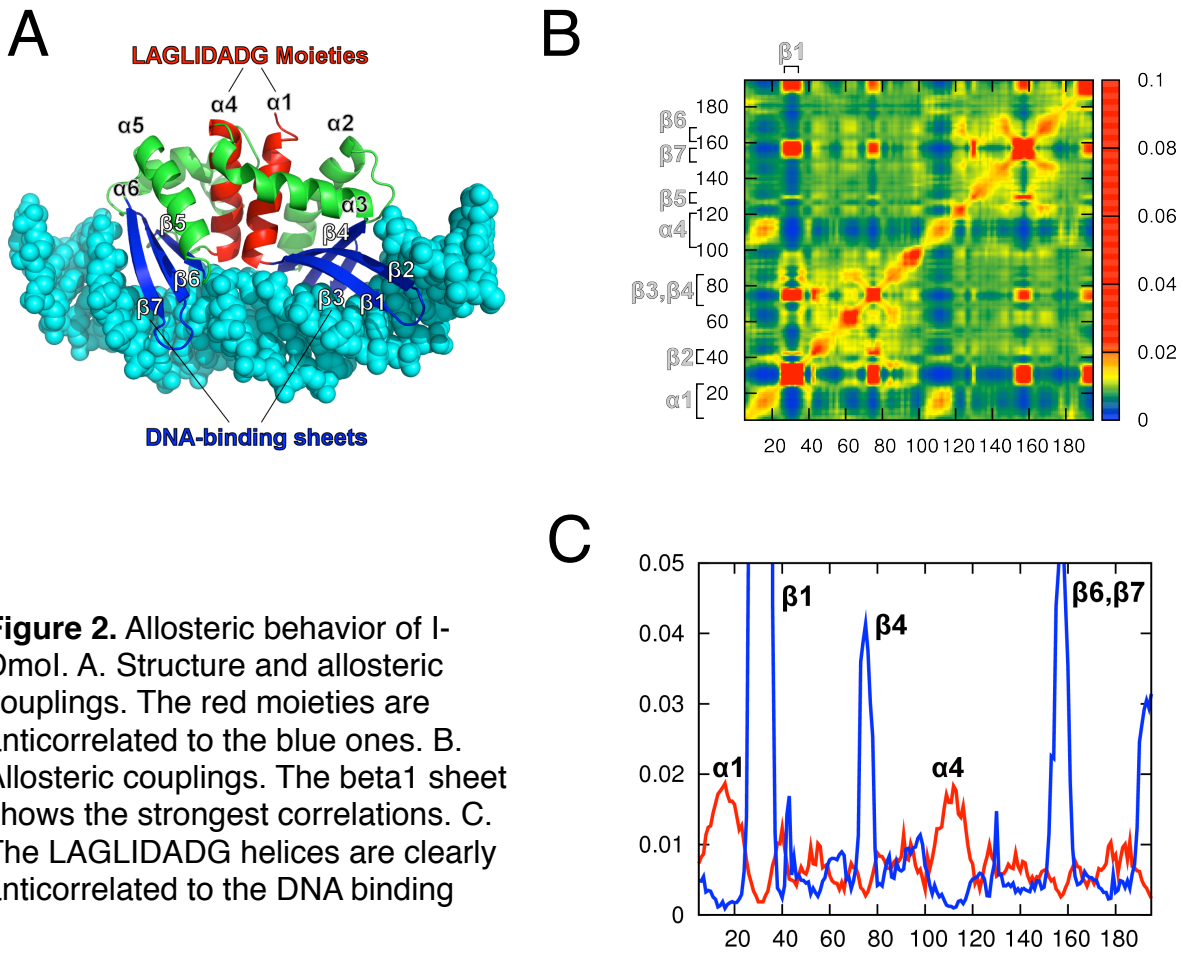

Figure 2. Allosteric behavior of I-DmoI. A. Structure and allosteric couplings. The red moieties are anticorrelated to the blue ones. B. Allosteric couplings. The beta1 sheet shows the strongest correlations. C. The LAGLIDADG helices are clearly anticorrelated to the DNA binding