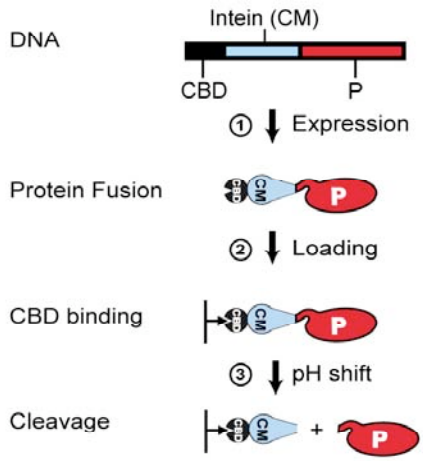
Inteins as Nanoswitches for Biotechnology: Linking Molecular Modeling with Physical and Genetic Methods

NSF NIRT Grant CTS03-04055

PIs: **Georges Belfort[1], Marlene Belfort[2], Victoria Derbyshire[2]**
Shekhar Garde[1], Saroj Nayak[1]
[1]Departments of Chemical and Biological Engineering and Physics, Rensselaer Polytechnic Institute and [2]Wadsworth Center, New York State Department of Health

Proteins perform an extraordinary array of functions within and outside living cells, with great specificity and efficiency. Certain proteins, called enzymes, can catalyze chemical reactions (i.e. speed up the conversion of reactants to products). Improving the catalytic properties of bio-molecular machines such as natural proteins or designing entirely new proteins through molecular engineering (using rational design and directed evolution) is likely to play an important role in the future of the biotechnology industry. These approaches are also expected to allow the design of novel nano-devices for biomedical applications for personalized medicine. In this work, "inteins", a particularly unique class of enzymes, will be harnessed to develop highly efficient single-step bioseparation processes as well as to design sensitive nanoswitches that could be used in sensors and other nano-devices for drug release in the treatment of diseases [1]. Inteins are 'proteins-within-proteins' and are inserted within the sequence of a "host" protein. They have the unusual property of being able to "cut" themselves out of their protein host and join the two remaining ends of the host protein (this process is called *splicing* while the process of just "cutting or bond breaking" is called *cleavage*) [2,3]. Although they occur naturally, they can be inserted into or combined with other valuable proteins to effect many different and very useful applications. One such application involves the use of inteins for improving manufacturing and purification of valuable new pharmaceutical products [4] (**Fig. 1**).

Fig. 1 Intein-mediated purification system.



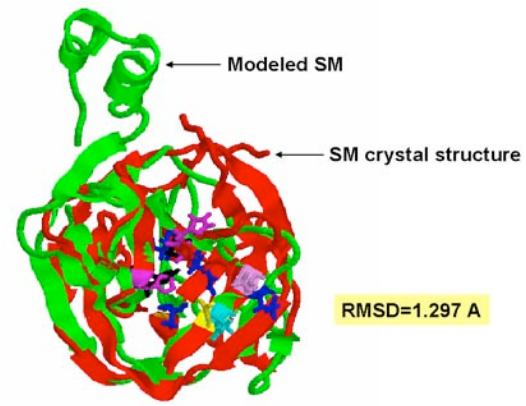
Strategy for purification of a target protein (P) using controllable CM intein (I) cleavage with a chitin-binding domain (CBD) in a tripartite fusion protein. Step 1. Expression of the tripartite fusion precursor protein *in vivo*. Step 2. After cell lysis and clarification in chitin column buffer (pH 8.5), the cell lysate is loaded onto a chitin column. Step 3. After washing the column, the pH is shifted from 8.5 to 6.0 to trigger C-terminal cleavage and the

The goals of this four-year cross-disciplinary, multi-investigator, multi-institutional research project are to determine the underlying principles of the splicing and cleavage reactions that occur during protein processing and to use this understanding to design a molecular nanoswitch that exhibits desirable properties for use in functional genomics and proteomics. In this, the first year, we have focused on molecular structure to gain insight into the cleavage and splicing processes of an autocatalytic self-processing mini-intein (mutants SM and CM from the *Mtu RecA* intein molecule). Later in the grant's lifetime, we will adapt the intein as a nanoswitch. Three complementary synergistic

approaches are being pursued: (i) A novel two-prong **theoretical approach** involving classical molecular dynamics and quantum *ab initio* calculations is used to unravel the mechanism of intein cleavage, (ii) **genetic approaches** to search for improved characteristics such as faster cleavage rate, reduced size, and different triggers for cleavage, and (iii) **modern biophysical methods** to follow the global and secondary structural changes occurring within the intein during cleavage and splicing. Research progress in each approach is described below:

(i) Theoretical approach. Initially, since X-ray or NMR structural data on our SM and CM mutant intein molecules were lacking, we performed homology based modeling of the SM mutant. Of the three intein structures available, in the best case (that of Mxe GyrA intein), the sequence identity was only about 30% which made the modeling of the three dimensional structure for the SM mutant difficult. Hence, a combination of molecular modeling and molecular dynamics simulation was used to obtain an estimate structure of the SM mutant intein [5]. Specifically, beginning with the Mxe GyrA template, we sequentially changed amino acid residues (2-3 mutation per step) to convert it into the SM mutant sequence over about 50 total steps.

Fig. 2. Comparison of the molecular model and crystal structure of Mtu RecA intein (SM) (From P. Van Roey).



Active site residues are shown in sticks, and the root mean square distance (RMSD) is that of those residues.

Recently, Patrick Van Roey, one of our collaborators, has obtained the crystal structures of the SM and CM mutant inteins, which has allowed comparisons of our predicted structure with his structure. We find that the locations of active site residues (which are also conserved in many canonical inteins) as well as the overall horseshoe-like structure of the protein were well reproduced in our predictions (**Fig. 2**).

The availability of the X-ray structure of the SM mutant intein has also allowed us to perform detailed molecular dynamics simulations in water as a function of pH and recently to investigate the mechanism of C-terminal cleavage. To mimic high and low values of pH we have performed simulations with histidine residues in both charged and uncharged states. We have also performed simulations of both splicing (SM) and cleaving (CM) mutants to investigate the structural as well as dynamic differences between the two structures.

To understand the mechanistic and energetic aspects of chemical reactions in inteins, we used a combination of *ab initio* quantum mechanical calculations (allows bond breaking) along with classical molecular mechanics (provides the dynamics and locations of atoms) focusing on the specifics of C-terminal cleavage process. For QM/MM calculations we focused on optimization of structures and their energies using a two-layer system (QM/MM). The outer layer consisted of the MM region and the inner layer was treated using QM techniques. C-terminal cleavage is initiated by the attack of the nitrogen of the amide of aspargine on the carbonyl carbon of the sessile peptide bond. We explored the structure local to the ASN residue with specific emphasis on the following three requirements: (i) a protein residue or water molecules needs to abstract the proton from ASN amide to make it nucleophilic (ii) protein residue or water molecules are required to be positioned in the correct position to stabilize the

oxyanion intermediate, and (iii) a proton needs to be donated to the leaving amide of the cleaved peptide amide. We are currently exploring the extent to which the penultimate HIS residue as well as other vicinal HIS residues may play a role in fulfilling one or more of the above three key requirements. These and similar studies will not only continue to provide mechanistic understanding but will also provide interpretation of the mutation studies performed by the molecular biology team (Derbyshire and M. Belfort).

(ii) **Genetic approaches:** As mentioned above, Patrick Van Roey (who is not funded by this grant yet has worked gratis for the grant and has made a major contribution), in collaboration with M. Belfort and V. Derbyshire, has solved the crystal structures of the SM and CM mini-inteins. While the crystallography data still needs final refinement, they have provided much information to bolster our confidence in the modeling and to further our continued development of smaller and more efficient controllable inteins. Preliminary comparisons of these crystallographic structures with the calculated ones (see above) and with those from the literature show they adopt the same common fold, as we would expect (**Fig. 2**).

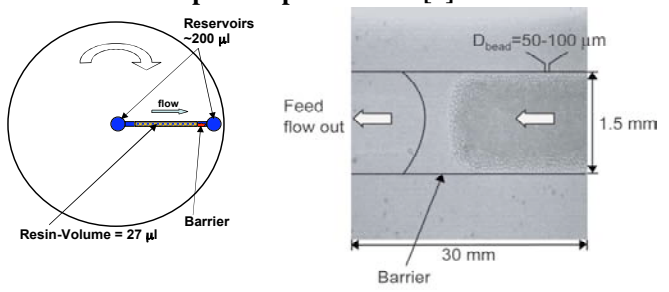
The first goal of the genetic effort has been to generate smaller and more active mini-intein derivatives based on the new structural information we have available. The mini-intein derivatives were generated by post-doctoral fellow, Kaori Hiraga, who deleted the residual endonuclease domain of the 168-amino acid min-intein. This work using multiple sequence alignment showed which amino acid strings could or could not be deleted. We thus used directed mutagenesis, in combination with genetic screening schemes, to derive active, stable, pH-controllable mini-inteins from a 168-amino acid parent. A number of derivatives have been generated and screened for *in vivo* splicing activity using the thymidylate synthase (TS) reporter system. These range in size from 135-139 amino acids. We plan to investigate the relative stability of these derivatives using biophysical methods and also carry out additional X-ray crystallographic analyses. All have been overexpressed and purified to homogeneity for these studies.

(iii) **Modern biophysical methods:** The fluorescence and mass spectrometry characterizations of the *Mtu* RecA mini intein are being performed through the development of spectroscopically suitable intein constructs at the Wadsworth Center and analyses performed at RPI by post-doctoral fellow, Tara M. Snyder. Ultimately, a more lucid understanding of the intein cleavage and splicing mechanisms is desired as a function of variable conditions such as pH and temperature. A fluorescent tripartite intein construct was designed with two fluorescent protein tags, cyan fluorescent protein (CFP) and yellow fluorescent protein (YFP) separated by the CM mutant intein. The emission spectra of CFP and the excitation spectrum of YFP overlap, allowing for Fluorescence Resonance Energy Transfer (FRET) between the fluorophores. Thus, FRET is being utilized to monitor intein cleavage and splicing reactions as a function of temperature and pH. Preliminary results suggest that both CFP and YFP were active fluorophores in the tripartite fusion construct. A novel mini-precursor construct (pET45b) has been designed with SM and CM inteins and is currently in the purification stage for the analysis by hydrogen exchange mass spectrometry. The constructs will be used to monitor amino acid-specific conformational changes as a function of the cleavage and splicing mechanisms.

For small-scale massively parallel applications on a fluidics platform, it is imperative that the purification process be as simple as possible with a minimum number of recovery steps and minimum addition of co-factors or other chemicals. As part of this grant, we have reported on a

single-step fusion-based affinity purification of proteins from *E.coli* lysates with pH-controllable linkers and a chitin binding domain (CBD) in a fluidic device [6]. Two different linkers were generated to solve two distinct separation problems: one (CM or cleaving mutant mini-intein) for rapid single-step affinity purification of a wide range of proteins, and the other (SM or splicing mutant mini-intein) specifically for the purification of cytotoxic proteins (**Fig. 3**).

Fig. 3. Fluidics channel on a rotating compact disk for intein-mediated protein purification [6].



(LEFT) Plan view of scaled-down fluidic channel on a rotating compact disk with PDMS covers and filled with affinity beads. Fluid flow is from the loading (feed or inner) reservoir to the eluting (product or outer) reservoir (arrow) and is driven by centrifugal force. The rotational speed is 250 RPM. (RIGHT) Optical micrograph of the fluidics channel with 50-100 µm diameter chitin coated

Educational Outreach Initiative

The collaborative research team orchestrated a biotechnology seminar and demonstration for the Questar III New Visions program (https://www.questar.org/); a public educational support agency for secondary schools in the Rensselaer, Columbia, and Greene counties of New York State. Under the local direction of Ms. Tammie Borland, the objective of Questar New Visions is to expose high school students to career paths in math, engineering, technology, and science. The lectures and demonstrations were attended by seniors from local high schools. Our objective was to expose the students to university research facilities in order to provide a wider appreciation for careers in the research sciences. Four seminar topics were presented: inteins, recovery of transgenic milk proteins, synthetic membrane technology, and proteins at interfaces. The session was very well-received, as indicated by favorable comments in exit questionnaires, describing the presentations as "fun" and "great explanation of difficult topics." All students reacted favorably to the laboratory tours (several demonstrations were set up), noting that the equipment and experiments were "state of the art," "…not things I see in normal classes," "…couldn't believe the price of some of the small pieces."

References

[1] For further information about this project email Georges Belfort <belfog@rpi.edu>, Marlene Belfort <Marlene.Belfort@wadsworth.org>, Victoria Derbyshire <vicky.derbyshire@wadsworth.org>, Shekhar Garde <gardes@rpi.edu>, and Saroj Nayak <nayaks@rpi.edu>.

[2] Belfort, M., Derbyshire, V., Stoddard, BL, and Wood, DW (Editors) (2004) Homing endonucleases and inteins, Springer-Verlag, New York, in press.

[3] Belfort, M., (2004) "Back to basics: structure, function, evolution and application of homing endonucleases and inteins", Chapter 1 in Homing endonucleases and inteins, (Belfort, M., Derbyshire, V., Stoddard, BL, and Wood, DW (Eds) Springer-Verlag, New York, in press.

[4] Wood, DW, Harcum, S., and Belfort, G (2004) "Industrial applications of inteins", Chapter 20 in Homing endonucleases and inteins, (Belfort, M., Derbyshire, V., Stoddard, BL, and Wood, DW (Eds) Springer-Verlag, New York, in press.

[5] Biswas, P, Storm, E., Belfort, G and Garde S (2004) "Molecular Modeling of Mtu Rec A Intein Structure", in preparation.

[6] Miao J., Wu, W, Spielmann, T., Belfort, M., Derbyshire V. and Belfort G. (2004) Single-step affinity purification of toxic and non-toxic proteins on a fluidics platform, *Lab-on-a-Chip*, submitted.