Chapter 42

Molecular Modeling of Fluorescent SERCA Biosensors

Bengt Svensson, Joseph M. Autry, and David D. Thomas

Abstract

Molecular modeling and simulation are useful tools in structural biology, allowing the formulation of functional hypotheses and interpretation of spectroscopy experiments. Here, we describe a method to construct in silico models of a fluorescent fusion protein construct, where a cyan fluorescent protein (CFP) is linked to the actuator domain of the Sarco/Endoplasmic Reticulum Ca^{2+}-ATPase (SERCA). This CFP-SERCA construct is a biosensor that can report on structural dynamics in the cytosolic headpiece of SERCA. Molecular modeling and FRET experiments allow us to generate new structural and mechanistic models that better describe the conformational landscape and regulation of SERCA. The methods described here can be applied to the creation of models for any fusion protein constructs and also describe the steps needed to simulate FRET results using molecular models.

Key words Molecular modeling, Fluorescent proteins, SERCA, Fluorescence resonance energy transfer, Biosensor

1 Introduction

Molecular modeling and simulations are useful tools for biophysical research. More than just creating pretty pictures for presentations and publications, modeling also assists greatly in the interpretation of experimental results. Fluorescence spectroscopy, including time-resolved anisotropy and FRET (Fluorescence Resonance Energy Transfer), is used in our laboratory to study structural dynamics of the proteins involved in Ca^{2+} transport and its regulation. The aim is to generate new structural and mechanistic models that better describe the conformational landscape and regulation of SERCA.

A biosensor is an analytical biomolecule that detects a molecular event such as ligand binding or protein-protein interaction. The biosensor field is developing rapidly, particularly in fluorescence methods, with new advances in fluorescence detection technologies in microplate instruments and biosensor engineering with genetically encoded probes such as green fluorescent protein

Maike Bublitz (ed.), *P-Type ATPases: Methods and Protocols*, Methods in Molecular Biology, vol. 1377,
DOI 10.1007/978-1-4939-3179-8_42,

(GFP) and its mutant color derivatives. We have used fluorescent fusion proteins of SERCA for FRET measurements, which were applied to study SERCA ensemble conformation, single-molecule structural dynamics, regulatory subunit interactions, and discovery of enzyme modulators via high-throughput screening [1–6]. Ion-motive ATPases have been tagged with GFP for cellular localization studies, including mammalian, worm, paramecium, plant, and parasite cells [7–12].

Here we describe a procedure to generate models and calculate fluorescence parameters of a fusion protein construct made up by cyan fluorescent protein (**CFP**) linked to the actuator domain of SERCA. X-ray crystal structures suggest that two cytosolic domains of SERCA, the nucleotide-binding and actuator domains, move apart by 3 nm upon Ca^{2+} binding, i.e., undergoing a calcium-induced transition from closed to open conformation of the cytosolic headpiece. We begin with building a fluorescent fusion protein model using the calcium-free (E2·Tg) state of SERCA (PDB ID: **1IWO**) [13]. The procedure to create molecular models for any other SERCA state would be similar; one just needs to select a different SERCA PDB file as starting point. The generated molecular models for the CFP-SERCA fusion proteins were used to predict distances between the fluorophores and fluorophore orientation of CFP-SERCA labeled with fluorescein isothiocyanate (FITC) in the nucleotide-binding domain. These data were used to calculate in silico FRET efficiency between CFP donor on the A domain with FITC acceptor on the N domain, and this predicted FRET efficiency was compared with experimental values measured in ER membranes in physiological solution. Together, experimental and computational results indicate that the time-average conformation of the SERCA headpiece is partly open in both the presence and absence of Ca^{2+}, i.e., not fully open or completely closed as suggested by X-ray crystallography [1]. These results were confirmed by single-molecule fluorescence of a "2-color" SERCA biosensor, indicating that SERCA undergoes a transition from an open, dynamic conformation to a closed, ordered structure and samples several discrete structural sub-states that are Ca^{2+} dependent [3]. The 2-color SERCA biosensor has proved to be a useful tool in identifying SERCA activators and inhibitors in high-throughput drug screens [6].

2 Materials

Molecular modeling, simulation, and molecular graphics software:

2.1 Discovery Studio Visualizer

Accelrys Discovery Studio® is a software suite of life science molecular design solutions for computational chemists and computational biologists. It is a molecular modeling environment for both small molecule and macromolecule applications, targeted mostly

towards the needs in drug discovery and pharmaceutical industry. DS Visualizer is a free version of the graphical interface to Discovery Studio, with some features missing compared with the commercial product. For example, most simulation tools have been removed. DS Visualizer is easy to use compared to most molecular graphics software packages and it has useful model building tools for working with small molecules and macromolecules. DS Visualizer is quite well documented and has several useful tutorials, tours, and sample files (available from the program Welcome screen) that will allow the user to get acquainted with the workings of the software quickly. DS Visualizer is available as a free download at http://www.accelrys.com. The current version is 4.1, but the modeling procedure described here works with any version from 3.1 to 4.1. Both Windows and Linux versions of the software are available. Here we describe the use of the Windows version of DS Visualizer, and the Linux version is essentially identical.

2.2 FPMOD

Fusion Protein MODeller (FPMOD), developed in the laboratory of Kevin Truong, University of Toronto, is a modeling software package that is designed to generate an ensemble of models of fusion proteins in order to sample the conformational space [14, 15]. This is done by treating the protein domains as rigid bodies and rotating each of them around the flexible linker connecting them. The FPMOD software can be obtained from the software authors at http://apel.ibbme.utoronto.ca/apel/software. The Windows version includes a graphical user interface (GUI) to the individual programs, which are located in the "scripts" folder. The software can be run from the GUI or the separate programs can be run from the command line. Instructions of how to use the software package can be found at the same website and in a publication by Pham and Truong [16].

2.3 Cygwin

Cygwin is a Unix-like environment and command-line interface for Microsoft Windows. The software package provides a large collection of GNU and Open Source tools which provide functionality similar to a Linux distribution on Windows. The Cygwin package can be downloaded from http://cygwin.com/. These tools are optional but convenient to process the data files generated from the modeling.

3 Methods

The modeling procedure for the CFP-SERCA fusion protein involves the following steps:

1. Customize the DS Visualizer program interface by showing Toolbars.
2. Load the amino acid sequence for the CFP-SERCA construct.

3. Load the protein structures from the PDB database, CFP (**1RM9**) and SERCA (**1IWO**).
4. Align the construct and structure sequences. Identify residues that are missing or different.
5. Build a molecular model of CFP (**1RM9**) by mutating mismatched residues, adding missing residues at N-terminus, and C-terminus.
6. Create a linker at the N-terminus of SERCA (**1IWO**).
7. Join CFP (**1RM9**) to the N-terminus of SERCA (**1IWO**).
8. Generate conformers by rotating backbone torsional angles in the linker region.
9. Save the models and create images.
10. Generate an ensemble of molecular models using FPMOD.
11. Extract fluorophore center and dipole vector coordinate position.
12. Calculate FRET parameters.

3.1 Customize the Program Interface by Showing Toolbars

Start the DS Visualizer program by clicking on the program icon on the computer desktop or on the "Discovery Studio Client" entry in the Start Menu. At the top below the top frame we have the menus as in most Windows programs, below that are the Tools tabs (starting with Macromolecules, Simulation etc.). Below these are Toolbars with commonly used functions. The work area is below these and consists of the Tools area to the left and tabbed display panels to the right (Fig. 1).

The interface is very customizable and can be set up to suit your preferences. For example, the Tools panel can be unpinned so that it is hidden, giving more space to the work area. The window can be undocked from the work area as well. There are help documents

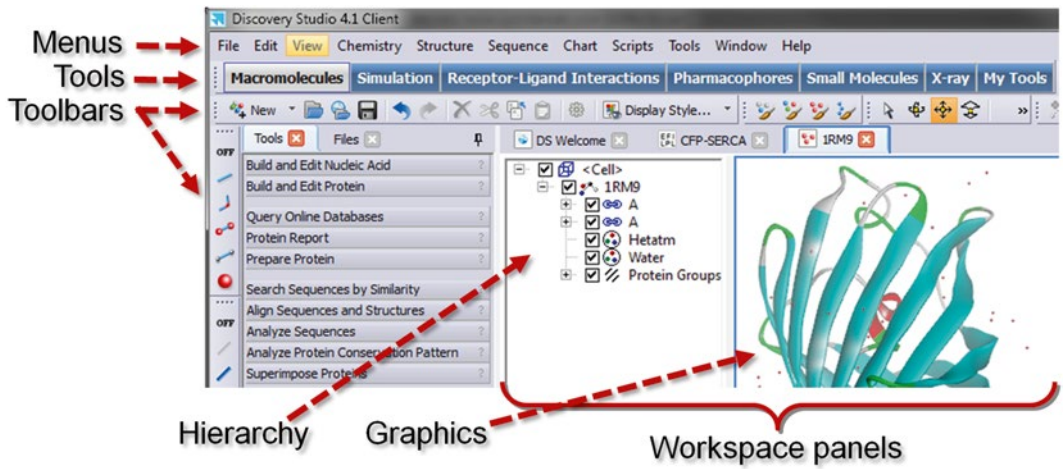

Fig. 1 Overview of the Discovery Studio Visualizer program interface

available on the "DS Welcome" screen or the "Help" menu that describe these features in more detail.

We want to add a few more Toolbars that have functions that we will use often. The easiest way to do that is to go to the "View" menu and go down to the "Toolbars" and when the next level pops up click on "Atom Display" if it does not have a check mark already. The new toolbar will now show above the Tools. The newly added "Toolbars" can be moved to any location. One useful choice is to move them to the left edge of the program window. Drag the toolbar to the left edge and release. Repeat these steps to also show the "ProteinStructure Display", "Sketching", and "Chemistry" toolbars. At this point it is recommended that some time is spent getting used to the program interface. The "Quick Start" tutorials, which can be found on the "DS Welcome" screen, are a good start for anyone unfamiliar with this software.

3.2 Load the Amino Acid Sequence for the CFP-SERCA Construct

To start the modeling procedure, first we want to load the protein sequence for the CFP-SERCA construct. Go to the "File" menu, and to "New" and to "Protein Sequence Window". This opens a new tab in the work area, with a blank sequence section. Copy the sequence data from Table 1 into the sequence section. Save the sequence data by going to the "File" menu and select "Save As…". Save the file as "CFP-SERCA.fa" in an appropriate location on your computer.

Alternatively, the sequence data can be downloaded from our website. Go to the "File" menu, and to "Open URL". In the top text box, type "http://www.msi.umn.edu/~bsven/CFP-SERCA.fa". Click the Open button. A new tab in the work area should appear with the protein sequence.

3.3 Load the Protein Structures from the PDB Database

Now we want to load the CFP and SERCA coordinate files from the Protein Data Bank [17]. We choose the **1RM9** crystal structure [18] as that one has the highest sequence similarity to the CFP protein of our fusion protein construct. Go to the "File" menu and to "Open URL". Type "**1RM9**", in the ID box and click the Open button. A new tab titled **1RM9** should have opened in the work area, where the CFP protein is shown in ribbon representation. Repeat this for the **1IWO** PDB ID, which is the atomic coordinate file for the structure of SERCA in the E2·Tg state, in which the inhibitor thapsigargin (Tg) locks SERCA in a calcium-free state with closed cytosolic headpiece [13]. Switch back to the **1RM9** tab by clicking on it. On the right side a ribbon representation of the proteins should be visible. On the left side of the work area the "Hierarchy" panel, this shows an overview of what is present in the PDB file (Fig. 1). If the "Hierarchy" panel is not shown already, use the "View" menu to toggle it on. The **1RM9** structure has waters in a second "**A**" chain. We want to delete all atoms except the first protein chain. Select the second "**A**" chain in the "Hierarchy" panel by clicking on it. It should be

Table 1
Sequence for the CFP-SERCA fusion protein construct

VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTL
VTTLTWGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV
NRIELKGIDFKEDGNILGHKLEYNYISHNVYITADKQKNGIKANFKIRHNIEDGSVQLAD
HYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYK*GE*
*L*MEAAHSKSTEECLAYFGVSETTGLTPDQVKRHLEKYGHNELPAEEGKSLWELVIEQFED
LLVRILLLAACISFVLAWFEEGEETITAFVEPFVILLILIANAIVGVWQERNAENAIEAL
KEYEPEMGKVYRADRKSVQRIKARDIVPGDIVEVAVGDKVPADIRILSIKSTTLRVDQSI
LTGESVSVIKHTEPVPDPRAVNQDKKNMLFSGTNIAAGKALGIVATTGVSTEIGKIRDQM
AATEQDKTPLQQKLDEFGEQLSKVISLICVAVWLINIGHFNDPVHGGSWIRGAIYYFKIA
VALAVAAIPEGLPAVITTCLALGTRRMAKKNAIVRSLPSVETLGCTSVICSDKTGTLTTN
QMSVCKMFIIDKVDGDFCSLNEFSITGSTYAPEGEVLKNDKPIRSGQFDGLVELATICAL
CNDSSLDFNETKGVYEKVGEATETALTTLVEKMNVFNTEVRNLSKVERANACNSVIRQLM
KKEFTLEFSRDRKSMSVYCSPAKSSRAAVGNKMFVKGAPEGVIDRCNYVRVGTTRVPMTG
PVKEKILSVIKEWGTGRDTLRCLALATRDTPPKREEMVLDDSSRFMEYETDLTFVGVVGM
LDPPRKEVMGSIQLCRDAGIRVIMITGDNKGTAIAICRRIGIFGENEEVADRAYTGREFD
DLPLAEQREACRRACCFARVEPSHKSKIVEYLQSYDEITAMTGDGVNDAPALKKAEIGIA
MGSGTAVAKTASEMVLADDNFSTIVAAVEEGRAIYNNMKQFIRYLISSNVGEVVCIFLTA
ALGLPEALIPVQLLWVNLVTDGLPATALGFNPPDLDIMDRPPRSPKEPLISGWLFFRYMA
IGGYVGAATVGAAAWWFMYAEDGPGVTYHQLTHFMQCTEDHPHFEGLDCEIFEAPEPMTM
ALSVLVTIEMCNALNSLSENQSLMRMPPWVNIWLLGSICLSMSLHFLILYVDPLPMIFKL
KALDLTQWLMVLKISLPVIGLDEILKFIARNYLEG

The CFP part in the sequence, residues 1 through 238, is underlined. The three residues, 239–241 added by sub-cloning, are shown in italics. The SERCA sequence is residues 242–1335

highlighted yellow when selected. Press the Delete key to delete the waters.

Click on the **1IWO** tab to bring that panel to the front. The **1IWO** structure has two SERCA protein molecules in it. We want to delete the second one, i.e., chain "**B**". There is also an inhibitor, thapsigargin, which is bound to the protein in the structure. We want to delete this molecule as well. Select and delete the second "**A**" and "**B**" chains.

3.4 Align the Construct and Structure Sequences

Next we want to align the sequences from the protein structures with the amino acid sequence for the fusion protein construct. The protein structure in PDB needs to be updated for the amino acid sequence of the expressed protein to fill in gaps in the crystal structure of CFP and to match the cDNA variant of CFP.

Click on the **1RM9** tab to bring the panel to the front. Go to the "Sequence" menu and click on "Show Sequence". A new tab with the **1RM9** sequence should be shown. Select the sequence by clicking on the text **1RM9** in the left panel. The sequence should be highlighted black. Copy the sequence by right clicking in the sequence in the right panel and selecting "Copy" from the pop-up menu.

Bring the CFP-SERCA sequence window to front by clicking on that tab. Right click in the right panel and select "Paste" to paste the **1RM9** sequence in with the **1RM9-1IWO** sequence. We can now close the **1RM9** sequence tab.

In the CFP-SERCA tab we will now manually align the sequences by adding two spaces in the beginning of the **1RM9** sequence. Click before the first residue in the sequence to place the cursor at this location. Press the space bar twice to insert spaces. The beginning part of the sequence alignment now shows in a darker color, which means the sequences match.

At position 65 we notice that the sequences are not aligned any more. That is because the CFP fluorophore is made up by residues Thr65, Trp66, and Gly67. The fluorophore is represented in the sequence by one X character. To align the rest of the sequence, click after the X at position 65 to place the cursor there and press the space bar twice to insert two blanks.

Next we want to make sure that when we select a residue in the sequence panel that the residue gets selected in the structure. To do this, go to the "Sequence" menu and select "Link Sequence and Structure". Make sure "**1RM9**" is selected in the top pull down widget and then click Link. Click and drag the sequence tab "CFP-SERCA" to the bottom half of the work area so that we can see the sequence data and the structures at the same time.

3.5 Build a Molecular Model of CFP

We want to change the **1RM9** structure to match the fusion protein construct sequence. Make sure that the "Macromolecule" tools are selected. Pin the tool panel open if needed. Click on "Build and Edit Protein" to show that tool's controls. Set the "Build Action" to "Mutate". Currently only the ribbon representation of the protein is shown. To be able to view and select specific atoms, we want to display all atoms of the protein with the "Line" display style. Click on the "**A**" chain in the "Hierarchy" panel to select the whole protein then click on the "Line" representation in the "Atom Display" tool panel.

Residue 57 is a fluorinated Trp in the crystal structure and we want to change it to a regular Trp. In the sequence window for the "**1RM9**" sequence, select the X at position 57. The X represents an unknown or non-standard amino acid in the sequence. Click on the "**1RM9**" tab to activate the structure panel again. Alternatively, the residue to be mutated can be selected in the "Hierarchy" panel instead of in the sequence window. We want to display the atoms

of this selected residue differently from the rest of the protein. Click on the "Stick" display style in the "Atom Display" toolbar. Click on the "Center Structure" button in the "View" toolbar to display the residue in the center of the panel. Under "Choose Amino Acid" in "Build and Edit Protein" click on "Trp". The fluorinated Trp is now a regular Trp.

Next we will mutate residue 80 from an Arg to a Gln. This mutation is a natural variation that may vary depending on cDNA clone. Repeat the steps as was done for Trp57 to select and display residue 80. Click on "Gln" in the "Build and Edit Protein" tool to mutate the Arg to a Gln.

We can see that the conformation is not optimal because the Gln side chain is too close to other atoms. To adjust its conformation we will use the "Search Side-Chain Rotamers" tool. Make sure residue 80 is selected. In the list of "Macromolecule" tools click on "Search Side-Chain Rotamers" to open the rotamer tools. While residue 80 is selected click on "Set Active Residue" to get the rotamer library for a Gln. Click on rotamer #1 and we should see that the residue side chain changes its position. The steric clashes have been removed, so the side chain structure is now optimized.

The next residue to mutate is Ala206 which should be Lys. This mutation A206K prevents CFP and YFP from self-assembly into dimers and tetramers and is available in most commercially available cDNA clones, thereby preventing aberrant FRET induced by CFP-YFP oligomerization. Repeat the process of residue mutation and side chain rotamer optimization as was done for Arg80. In this case try rotamer #2, which produces a valid conformation with no steric clashes.

We also want to add the two missing residues to the N-terminus. Select the first residue, Lys3, in the "Hierarchy" panel. Click on "Build and Edit Protein". Set the "Build Action" to "Insert". Set "Conformation" to "Extended". Click on "Ser" under "Choose Amino Acid". We have now added a Ser to the N-terminus. Select the Ser2 residue and repeat the process to add Val1. Note that GFP and its mutant color derivatives contain a signal sequence that is cleaved co-translationally, and that the convention for sequence numbering is to start with +1 at remaining N-terminal residue, which is Val1.

We want to optimize the geometry for these two residues by performing an energy minimization. Multiple residue selections can be made by holding down the Ctrl key while clicking a residue. Select both Val1 and Ser2 then click on the "Clean Geometry" button on the "Chemistry" toolbar. We should see the atoms move. Click the "Clean Geometry" button again a few times until there is not much change anymore. We can go back to the sequence window and remove the extra spaces in the beginning of the **1RM9** sequence to make the sequences aligned again.

Finally we want to add 12 residues on the C-terminus of CFP. Nine residues at the C-terminus are not resolved in the crystal structure and an additional three residues that resulted from the subcloning of the two cDNA molecules needs to be added to the CFP structure. Select the last residue in the **1RM9** structure, i.e., Ile229. Center the structure. Click on "Build and Edit Protein". Set the "Build Action" to "Create/Grow Chain". Set "Conformation" to "Extended". Click on "Thr" under "Choose Amino Acid" to add it to the end.

Select the new last residue, i.e., Thr230. Repeat these steps to add the remainder of the sequence "Leu-Gly-Met-Asp-Glu-Leu-Tyr-Lys-Gly-Glu-Leu" to the end of **1RM9**. If you make a mistake and add the wrong residue you can use the Mutate function to correct it.

We have now built the correct CFP and linker regions for the CFP-SERCA construct At this point we should save the modified **1RM9** coordinates. Go to the "File" menu, select "Save As". Set the file type to "Protein Databank Files". Name the file "**1RM9+Linker.pdb**". Select an appropriate location to save and click the Save button.

3.6 Add an N-Terminal Linker to SERCA

Next we need to prepare the **1IWO** structure so that we can attach the **1RM9+Linker** to the N-terminus of SERCA. We could align the sequences as we did for **1RM9** and the fusion protein, but this is not necessary in this case as the sequence does match exactly. We need to add a couple of residues to the N-terminus so that we have an overlapping region between the two proteins. We will add the last three residues, Gly-Glu-Leu, of the linker region to the beginning of the **1IWO** structure.

Display the **1IWO** structure with "Lines" only. Select residue Met1 and center the structure. Click on "Build and Edit Protein". Set the "Build Action" to "Insert". Set "Conformation" to "Extended". Click on "Leu" under "Choose Amino Acid". We have now added a Leu to the N-terminus but we have a problem. The residue extends into the protein. We need make some manual adjustments to turn the N-terminus towards the surface of the protein.

We can adjust the backbone dihedral angles to fix this. Click on the "Torsion" button in the "Sketch" toolbar. With this tool we can adjust the backbone dihedral with the mouse. Click on the bond between the C and CA atoms of residue Glu2. With the left mouse button pressed down slide the mouse left to right to adjust the dihedral. Set this value to about 50°. Next click on the Glu2 CA to N bond and adjust the dihedral to about –160°. Next click on the Met1 C to CA bond and adjust the dihedral to about –145°. The N-terminus should now be facing away from the protein.

Select residues Leu0, Met1, and Glu2. By holding down the Ctrl key while selecting one can add to or remove from a selection.

Click on the "Clean Geometry" button to optimize the geometry for these residues.

We can now continue adding a Glu and Gly to the N-terminus using the "Build and Edit Protein" tools in the same way as was done for Leu0. We can now save the modified 1IWO structure. A suggested filename would be "**1IWO+Linker.pdb**".

3.7 Join CFP to SERCA

Now we will fuse the two structures. Bring the **1RM9+Linker** tab to front and then select the whole chain "**A**". Right click on it and select "Copy". Bring the **1IWO+Linker** tab to front. Right click in molecule panel and select "Paste" to copy the **1RM9+Linker** atoms into the same window.

We now want to superimpose the overlapping regions of the linkers. In the "Hierarchy" panel select the first three residues of **1IWO** and select the last three residues of **1RM9+Linker**. Click on the "Superimpose Protein" in the "Macromolecules" tools to open that tools controls. Select "Tethers" in the "Superimpose by" pull-down list, then click on "Create Protein Tether". If this worked we will see lines from one protein to the other in the molecule panel. Click on "Superimpose", which can be found just below "Create Protein Tether". The overlapping sequences are now superimposed.

Center the view on the three consecutive residues Gly-2, Glu-1, and Leu0 of **1IWO**. Now we want to delete the extra residues. Select residues Gly-2 and Glu-1 of **1IWO** and press the delete key. Select Leu241 of **1RM9** and press the delete key.

To join the two proteins together we want to create a bond between the C atom of residue Glu240 of **1RM9** and the N atom of Leu0 of **1IWO**. Select only these two atoms in the graphics window by first clicking on one atom then holding down the shift key while clicking on the second atom. The atoms can also be selected in the "Hierarchy" panel by clicking one atom then clicking the second while holding down the Ctrl key. Once the two atoms are selected, click on the "Single Bond" button in the "Chemistry" toolbar.

Some tidying up of the chain assignment and residue numbers is recommended. There are now two "**A**" chains as seen in the "Hierarchy" panel. We can fuse these two chains into one by selecting the second "**A**" chain and dragging and dropping it over the first "**A**" chain in the "Hierarchy" panel.

The sequence numbering now goes from 1 to 240 for the **CFP+Linker** part and the last residue of the linker is Leu0 after which SERCA starts at Met1. To be able to do any further simulations on this model we need to correct the residue numbering of the CFP-SERCA construct so that the residue numbers are sequential. Select the residue Leu0 to Gly994 in the "Hierarchy" panel, and then open the "Prepare Protein" tool under the "Macromolecules" tools. Click on "Renumber Sequences…" and select "Renumber

residue range from Leu0 to Gly994". Make sure that the "Starting from residue ID" is set to 241 and click OK. The model for the CFP-SERCA fusion protein construct is now correctly built. Save it as "**1RM9+1IWO.pdb**".

3.8 Generate Linker Conformations

This model for the fusion protein with the linker region completely straight does not look very realistic (Fig. 2). In reality the linker region would likely have a much more random conformation. We can apply a few more modeling procedures in DS Visualizer that will generate more realistic conformations of the linker region. The "Build and Edit Protein" has a function to change the backbone conformation.

Open up that tool again and set "Conformation" to "Right-hand Alpha Helix". For example, select residues Thr230 through Leu241. Then click on "Apply Conformation". We can randomize it a little bit more by selecting one or two residues in the region between Thr230 and Leu241 and changing the backbone conformation to any other secondary structure setting.

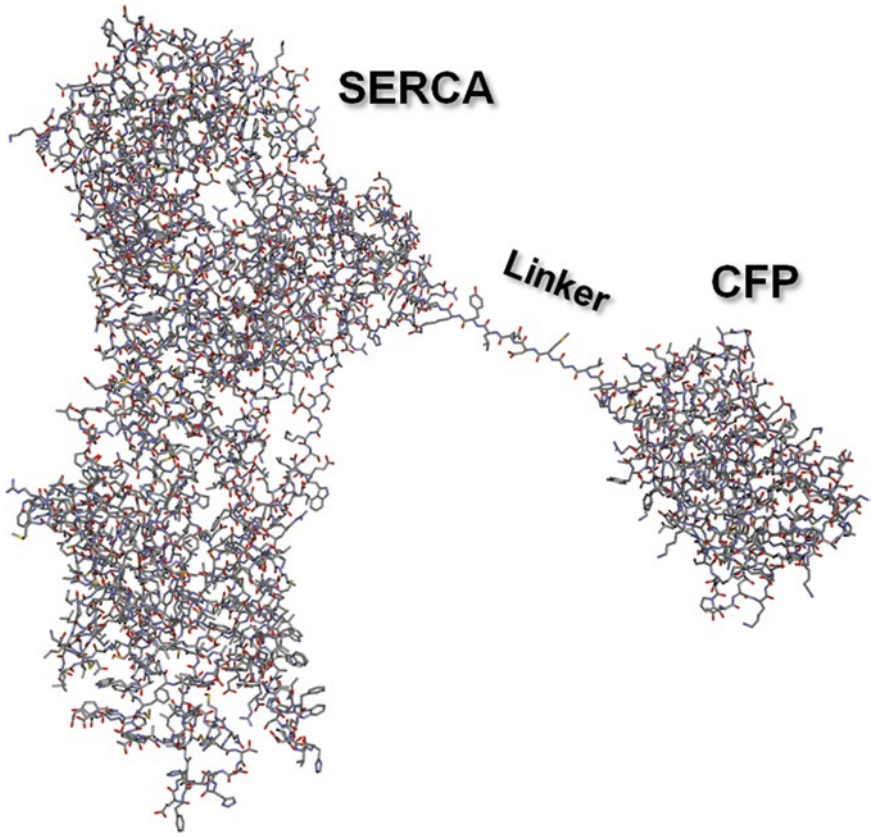

Fig. 2 Initial complete model for the CFP-SERCA fusion protein construct. Note the straight extended conformation of the linker region

Another way of generating conformations of the linker region is by using the "Torsion" tool in the "Sketching" toolbar to rotate dihedrals in the linker. Be careful to only rotate around N-CA or CA-C bonds.

If the CFP clashes with SERCA, i.e., has atom overlaps, one needs to try different dihedral values in the linker. When a conformation has been found that does not have any clashes, then we can select the whole linker region Thr230 trough Leu241 and use the "Clean Geometry" tool to optimize the structure of the selected region. This minimizes the energy by allowing atoms in both backbone and side chains to move. The linker region should now have a reasonable structure.

The model is now finished and we should save it. A suggested filename would be "**1RM9+1IWO_conf1.pdb**". Generate a few more random conformations for the linker region and save these as well. The model is good enough to use as starting point for simulations or to show in a figure for a presentation.

3.9 Making Publication-Quality Figures

DS Visualizer has the ability to save the molecular graphics view as an image and also the ability to export the graphics view as a POV-Ray scene for ray-tracing with the POV-Ray ray-tracing software, http://www.povray.org/. The molecular graphics view in DS Visualizer is quite good, although other software packages like VMD [19], http://www.ks.uiuc.edu/Research/vmd/, or PyMOL, https://www.pymol.org/, can produce somewhat nicer looking images. PDB format files created by DS Visualizer can be opened and visualized in either of these other molecular graphics software.

Choose a useful view and display it with an appropriate graphics representation, such as ribbons to better show the global protein structure (Fig. 3). Different sections of the protein can be displayed with different colors as shown here where the SERCA domains are shown in red, green, blue, and grey, and the CFP part of the construct is shown in cyan. When you are satisfied with how the display looks go to the "File" menu and select "Save As". Set the "Files of type" to "Image Files", and save the file as "CFP-SERCA.png". A requester will pop up asking what resolution the image should be saved at. Keep the current resolution or select a higher resolution for publication quality images.

Previously, we have saved the created models as PDB format files, but if we want to save both the view setting and the loaded molecule data we should save the data using the "Discovery Studio Files (*.dsv)" format. Save the file as "**1RM9+1IWO.dsv**". When loading the dsv file at a later time all view and display setting should be preserved.

3.10 Generate an Ensemble of Molecular Models

It is assumed that CFP can adopt many possible conformations around SERCA. To simulate the multitude of possible conformations that CFP can adopt with respect to SERCA in solution, we

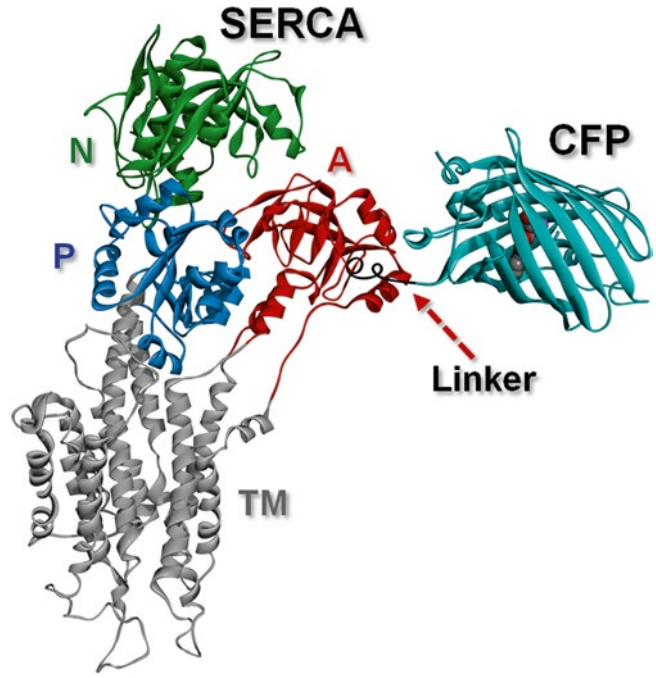

Fig. 3 A ribbon representation of the CFP-SERCA fusion protein construct is shown as an example of a figure suitable for structural analysis. SERCA is shown with the actuator domains in *red*, the nucleotide domain in *green*, the phosphorylation domain in *blue*, and the transmembrane domain in *grey*. The CFP part of the fusion protein is shown in *cyan*

need to create an ensemble of structural models. One way to do that is using the software Fusion Protein MODeller (FPMOD) [14, 15]. In short, SERCA and CFP structures are treated as rigid bodies joined by a flexible linker region that is allowed to rotate freely. The nine C-terminal residues of CFP, which are unresolved and assumed to be flexible in the crystal structure, and the residues of the Gly-Glu-Leu sequence at the N-terminus of SERCA are assigned to be part of this flexible linker. An ensemble of more than 1000 models needs to be generated to sample the conformational space of CFP around SERCA.

The easiest way to generate an ensemble of models using FPMOD is to run the software from the command line. Download the FPMOD software package from http://apel.ibbme.utoronto.ca/apel/software/FPModGUI_v2.rar. Unpack the archive to a location of your choice. Copy the file "conformation_sample_finals.exe" from the "scripts" folder to the location where you saved the PDB format coordinate file for the CFP-SERCA fusion construct. Open a "Command Prompt" from the Windows Start Menu. Change folder to the location where the coordinate files

were saved by using the "cd" command and the folder path to where you have saved the PDB files.

Next we need to make a change to the PDB file that will allow FPMOD to read all atoms. Open the file "RM9+1IWO_conf1.pdb" in a text editor, for example Notepad. Search for the text string "HETATM" and replace that with "ATOM" at all placed where it is found in the file. Make sure that there are two spaces after "ATOM", otherwise the columns in the file would get shifted and the file cannot be read correctly. Save the file. It can be saved using the same filename.

To run the conformational sampling and generate an ensemble of models type or copy the following command line into the "Command Prompt" window:

```
"conformation_sample_finals.exe 1RM9+1IWO_conf1.pdb 1RM9+
   1IWO_conf1_ 1000 1 A N AGITLGMDELYKGELMEA TLGMDELYKGEL"
```

After the program name in the command line above comes first the input filename, which is the file we created in the previous steps. Next is the base name for the output filename, which we chose here as the input file name without the ".pdb" extension but with an underscore character added to the end. The output file will be saved with a model number and the extension ".pdb" at the end. After that we have the number of models to generate, in this case it is "1000". To get adequate sampling of the fusion protein conformation at least 1000 models needs to be generated. The next value on the command line is the number of linkers in our construct; in this case that is "1". Next is the chain ID, which is "A" in our construct. After that the command line lists either "N" or "C" depending on whether the fluorescent protein is attached to the N-terminal or C-terminal end of our protein. For our construct CFP is attached at the N-terminus, therefore an "N" is put in the command line. At the end of the command line there are two stretches of the sequence around the flexible linker in our construct. The first sequence has a few extra residues at each end compared to the second sequence. The second sequence comprises the residues that we select to be part of a flexible linker. The flexible region is assumed to be starting after the end of the last β-strand of CFP and end at the residue before the start of the SERCA sequence. This is the same sequence stretch of 12 residues that we modeled in at the C-terminus of CFP. The generation of thousands of models could take up to a couple of hours to complete.

A suggestion to reduce the time in half for generating 1000 models is to start multiple instances of the "conformation_sample_finals.exe" program in separate "Command Prompt" windows. For example start a first run of "conformation_sample_finals.exe" generating 500 models and then start another instance of the program also generating 500 models. A different starting conforma-

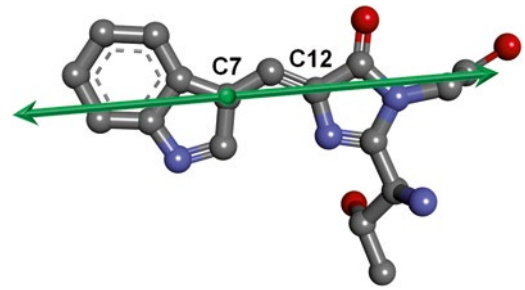

Fig. 4 The CFP fluorophore generated from the sequence Thr65-Trp66-Gly67. The vector centered on the C7 atom and going through the C12 atom represents the excitation and emission transition dipole (*green arrow*)

tion is needed for every run; otherwise the same conformations will be generated. The starting conformation can either be another manually generated conformation or it could also be any of the models created by the first run of the program. Remember to give each output file a different base name so that you will not overwrite the previous data.

3.11 Extracting Fluorophore Center and Dipole Vector Coordinate Positions

For each structure in the ensemble, we can calculate both the distance between the centers of the two fluorophores (R) and the Förster distance (R_0), which is dependent on the relative orientation of the fluorophores. One property that needs to be known is the transition dipole moment for the fluorophore. Transition dipole moments for fluorescent proteins have been determined experimentally [20] and from excited state quantum chemistry calculations [21]. The transition dipole can be approximated quite accurately in the CFP fluorophore by setting the fluorophore center at the C7 atom and the dipole in the direction of the C12 atom (Fig. 4).

We want to extract the X, Y, and Z coordinate values for these atoms from all 1000 pdb files generated in the previous step for further calculations of FRET parameters. One could open every file in a text editor and copy and paste the data, but that would be very tedious work. A better option is to write a program or script to do that and to also calculate the FRET parameters described below. Here we show a simple way to extract the data from the pdb files using Unix command line tools. To get access to these tools one would need a computer running Linux or one could install the Cygwin Unix-like environment for Windows.

The commands below will extract the X, Y, and Z coordinates from all files in the folder that match the "1RM9+1IWO_conf1_*.pdb" pattern, which should be only the models generated by

FPMOD. The final command merges all these temporary files, "x1.txt" to "z2.txt", into a new file "xyz12.txt", which is a Tab delimited file that can be opened in Excel or any other spreadsheet or scientific graphing and data analysis software.

```
grep -h 'C7  4F3  A 66' 1RM9+1IWO_conf1_*.pdb | cut -c
   31-38 > x1.txt
grep -h 'C7  4F3 A  66' 1RM9+1IWO_conf1_*.pdb | cut -c
   39-46 > y1.txt
grep -h 'C7  4F3 A  66' 1RM9+1IWO_conf1_*.pdb | cut -c
   47-54 > z1.txt
grep -h 'C12 4F3 A  66' 1RM9+1IWO_conf1_*.pdb | cut -c
   31-38 > x 2.txt
grep -h 'C12 4F3 A  66' 1RM9+1IWO_conf1_*.pdb | cut -c
   39-46 > y2.txt
grep -h 'C12 4F3 A  66' 1RM9+1IWO_conf1_*.pdb | cut -c
   47-54 > z2.txt
paste x1.txt y1.txt z1.txt x 2.txt y2.txt z2.txt >
   xyz12.txt
```

The first three columns are the X, Y, and Z coordinates for the FRET donor fluorophore center, and the next three columns are X, Y, and Z coordinates representing the FRET donor transition dipole direction.

3.12 Calculating FRET Parameters

Molecular models of fusion protein constructs including both fluorescence donor and acceptor probes can be used to calculate expected FRET efficiency values. Here we have focused on modeling CFP as the donor fluorophore attached to the actuator domain at the N-terminus of SERCA. Similar modeling strategies can be used to add acceptor labeling for intramolecular or intermolecular FRET.

For example, acceptor labeling strategies for CFP-SERCA include intramolecular FRET studies using FITC, a small fluorescent probe that binds in the nucleotide binding pocket of SERCA [1], and for intermolecular FRET using YFP-tagged regulatory subunits phospholamban and sarcolipin [4, 5, 22, 23]. Another example is "2-color-SERCA", that was developed in the laboratories of Seth Robia, Loyola University Chicago, and David Thomas, University of Minnesota, in which GFP is fused to an interior loop of the nucleotide-binding domain and red fluorescent protein (RFP) is fused to the N-terminus in the actuator domain. This GFP-RFP-SERCA construct has been utilized for time-resolved detection of GFP-RFP intramolecular FRET in single-molecule microscopy and high-throughput drug discovery [2, 3, 6].

For the rest of this protocol, we will use a docked conformation of FITC by Winters et al. [1] as the FRET acceptor. The X, Y, and Z coordinates for the fluorophore center position and the

Table 2
Coordinates for FITC docked in the nucleotide pocket of the calcium-free (E2·Tg) structure of SERCA (PDB ID: 1IWO) [13]

	X	*Y*	*Z*
FITC center coordinates	−12.692	−18.062	58.889
FITC dipole vector	−13.679	−18.695	58.734

transition dipole vector are shown in Table 2. Copy these values to the spreadsheet. In the spreadsheet we can now calculate the FRET donor to acceptor distance, R, using

$$R = \sqrt{[(x_A - x_D)^2 + (y_A - y_D)^2 + (z_A - z_D)^2]} \quad (1)$$

where, x_D, y_D, and z_D are the donor center fluorophore coordinates and x_A, y_A, and z_A are the acceptor center fluorophore coordinates.

The fluorophore coordinate data can be used to calculate the Förster distance R_0. R_0 is dependent on the orientation factor κ^2, which can be calculated explicitly from the observed probe orientations using

$$\kappa^2 = (\sin\theta_D \sin\theta_A \cos\phi + 2\cos\theta_D \cos\theta_A)^2 \quad (2)$$

θ_D and θ_A are the angles between the inter-probe vector and the donor and acceptor transition moments, respectively, and ϕ is the dihedral angle between the two planes defined by the donor transition moments and the line connecting the centers of the donor and acceptor fluorophores (DR) and the acceptor transition moments and the line connecting the centers of the donor and acceptor fluorophores (AR) [24]. The θ_D and θ_A angles can be calculated from the coordinates using

$$\theta = \text{acos}\,[(A \bullet B)/|A||B|] \quad (3)$$

where A represents the fluorophore transition dipole vector and B represents the fluorophore–fluorophore vector. ϕ can be calculated using

$$\phi = \text{acos}\,([(A_2 - A_1)\times(A_3 - A_1)] \bullet [(B_2 - B_1)\times(B_3 - B_1)]/ |(A_2 - A_1)\times(A_3 - A_1)||(B_2 - B_1)\times(B_3 - B_1)|) \quad (4)$$

where A_1, A_2, and A_3 are three coordinate positions that describe the DR plane and B_1, B_2, and B_3 are three coordinate positions

that describe the AR plane. Functions to calculate the angles and dihedrals can be found on the web for most spreadsheet and data analysis software.

A κ^2 dependent R_0, $R_0(\kappa)$, can thus be calculated for each model from the simulations using

$$R_0(\kappa) = R_0(^2/_3) \times (3\kappa/2)^{1/3} \quad (5)$$

where $R_0(^2/_3)$ is the published R_0 value for the fluorescent probe pairs assuming random orientation of the probes. This value is 55 Å for in case of the CFP-FITC FRET pair [1]. The averaging regime should also be considered for FRET analysis [25, 26]. In the case of fluorescent fusion proteins the static averaging regime is appropriate as the rotational motion is slow for fluorescent proteins in relation to their fluorescence lifetime, which eliminates the need to average κ^2 over time before calculating the FRET efficiency. We can now calculate not only the predicted distribution of distances between the fluorophores (R), but also the predicted values of measured energy transfer efficiency (E) from an ensemble of conformation of fluorophores, using

$$E = 1 / [1 + (R / R_0(\kappa))^6] \quad (6)$$

Next we want to calculate the probability distribution for the distances between the fluorophores (R) and energy transfer efficiency (E) and make plots using graphing software of your choice. Origin, http://originlab.com/, was used to calculate the probability distribution of the data and to make the plots in Fig. 5.

The distance distribution, based on modeling and simulations, is shown to be quite broad. The full-width half-maximum (FWHM) of the distribution is approximately 30 Å (Fig. 5a). The orientation factor κ^2 is 0.6, which is slightly lower than what is predicted from a completely freely moving and isotropic case where κ^2 is $^2/_3$ (Fig. 5b). The orientation corrected $R_0(\kappa)$ also shows a broad distribution with an average of 51 Å, which is lower that the experimentally determined R_0 of 55 Å for the isotropic case (Fig. 5c). The simulated FRET efficiency distribution show a large population that show low FRET, less than 0.1. However, there are a significant number of conformations that show quite high FRET. The average is 0.290, which is close to the experimentally determined value of 0.336 for CFP-FITC-SERCA [1]. Winters et al. describe more details of this type of FRET analysis as used in the study of the SERCA cytosolic headpiece structural dynamics [1].

In conclusion, we have presented a method for modeling, conformational sampling simulations, and FRET parameter analysis of a fluorescent fusion protein construct of CFP and SERCA. The methods described would be generally applicable for the modeling and simulation of any other fusion protein construct independent of the proteins involved. We hope that this modeling method will be useful for the research community.

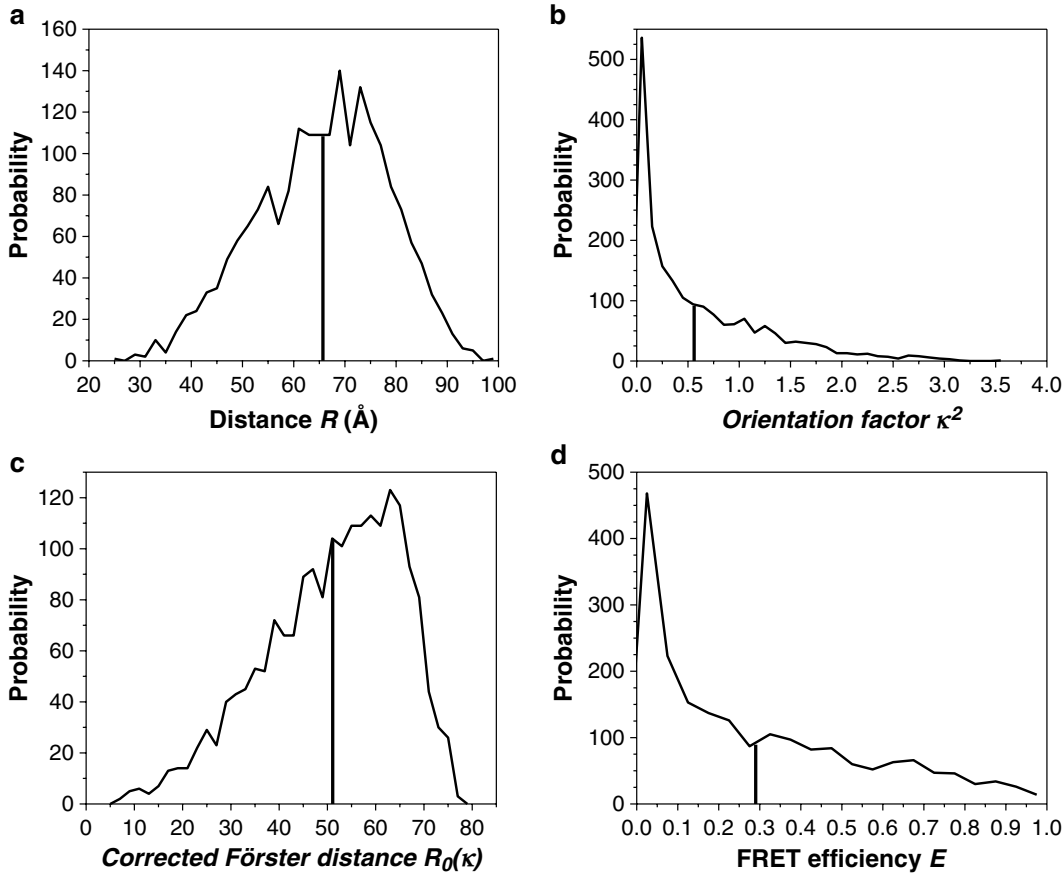

Fig. 5 Simulated FRET parameters based on molecular modeling of CFP-SERCA. Distributions were calculated for (**a**) distance, (**b**) κ^2, (**c**) κ^2-corrected R_0, and (**d**) FRET efficiency. A *vertical line* indicates the mean value for each calculated parameter. The mean value of *E* is comparable to the experimentally measured steady state FRET

Acknowledgement

This work was supported by NIH grants GM27906 and AR007612. Computational resources were provided by Minnesota Supercomputing Institute. This method chapter was previously presented as a hands-on tutorial at a regional Biophysical Society Networking Symposium at St. Olaf College, Northfield, MN.

References

1. Winters DL, Autry JM, Svensson B et al (2008) Interdomain fluorescence resonance energy transfer in SERCA probed by cyan-fluorescent protein fused to the actuator domain. Biochemistry 47:4246–4256
2. Hou Z, Hu Z, Blackwell DJ et al (2012) 2-Color calcium pump reveals closure of the cytoplasmic headpiece with calcium binding. PLoS One 7:e40369

3. Pallikkuth S, Blackwell DJ, Hu Z et al (2013) Phosphorylated phospholamban stabilizes a compact conformation of the cardiac calcium-ATPase. Biophys J 105:1812–1821
4. Autry JM, Rubin JE, Pietrini SD et al (2011) Oligomeric interactions of sarcolipin and the Ca-ATPase. J Biol Chem 286:31697–31706
5. Gruber SJ, Haydon S, Thomas DD (2012) Phospholamban mutants compete with wild type for SERCA binding in living cells. Biochem Biophys Res Commun 420:236–240
6. Gruber SJ, Cornea RL, Li J et al (2014) Discovery of enzyme modulators via high-throughput time-resolved FRET in living cells. J Biomol Screen 19:215–222
7. Vagin O, Denevich S, Sachs G (2003) Plasma membrane delivery of the gastric H, K-ATPase: the role of beta-subunit glycosylation. Am J Physiol Cell Physiol 285:C968–C976
8. Cho JH, Bandyopadhyay J, Lee J et al (2000) Two isoforms of sarco/endoplasmic reticulum calcium ATPase (SERCA) are essential in Caenorhabditis elegans. Gene 261:211–219
9. Hauser K, Pavlovic N, Klauke N et al (2000) Green fluorescent protein-tagged sarco(endo) plasmic reticulum Ca2+-ATPase overexpression in Paramecium cells: isoforms, subcellular localization, biogenesis of cortical calcium stores and functional aspects. Mol Microbiol 37:773–787
10. Gravot A, Lieutaud A, Verret F et al (2004) AtHMA3, a plant P1B-ATPase, functions as a Cd/Pb transporter in yeast. FEBS Lett 561: 22–28
11. Lefebvre B, Batoko H, Duby G et al (2004) Targeting of a Nicotiana plumbaginifolia H+-ATPase to the plasma membrane is not by default and requires cytosolic structural determinants. Plant Cell 16:1772–1789
12. Furuya T, Okura M, Ruiz FA et al (2001) TcSCA complements yeast mutants defective in Ca2+ pumps and encodes a Ca2+-ATPase that localizes to the endoplasmic reticulum of Trypanosoma cruzi. J Biol Chem 276:32437–32445
13. Toyoshima C, Nomura H (2002) Structural changes in the calcium pump accompanying the dissociation of calcium. Nature 418: 605–611
14. Pham E, Chiang J, Li I et al (2007) A computational tool for designing FRET protein biosensors by rigid-body sampling of their conformational space. Structure 15:515–523
15. Chiang J, Li I, Pham E et al (2006) FPMOD: a modeling tool for sampling the conformational space of fusion proteins. Conf Proc IEEE Eng Med Biol Soc 1:4111–4114
16. Pham E, Truong K (2010) Design of fluorescent fusion protein probes. Methods Mol Biol 591:69–91
17. Berman HM, Westbrook J, Feng Z et al (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242
18. Budisa N, Pal PP, Alefelder S et al (2004) Probing the role of tryptophans in Aequorea victoria green fluorescent proteins with an expanded genetic code. Biol Chem 385: 191–202
19. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14(33–38):27–38
20. Rosell FI, Boxer SG (2003) Polarized absorption spectra of green fluorescent protein single crystals: transition dipole moment directions. Biochemistry 42:177–183
21. Ansbacher T, Srivastava HK, Stein T et al (2012) Calculation of transition dipole moment in fluorescent proteins – towards efficient energy transfer. Phys Chem Chem Phys 14: 4109–4117
22. Kelly EM, Hou Z, Bossuyt J et al (2008) Phospholamban oligomerization, quaternary structure, and sarco(endo)plasmic reticulum calcium ATPase binding measured by fluorescence resonance energy transfer in living cells. J Biol Chem 283:12202–12211
23. Hou Z, Robia SL (2010) Relative affinity of calcium pump isoforms for phospholamban quantified by fluorescence resonance energy transfer. J Mol Biol 402:210–216
24. Dale RE, Eisinger J, Blumberg WE (1979) The orientational freedom of molecular probes. The orientation factor in intramolecular energy transfer. Biophys J 26:161–193
25. Van Der Meer BW, Coker Iii G, Chen S-YS (1994) Resonance energy transfer, theory and data. VCH, New York, NY
26. Vanbeek DB, Zwier MC, Shorb JM et al (2007) Fretting about FRET: correlation between kappa and R. Biophys J 92:4168–4178