Fwd: Structure Decision

1 message

Joel Rozowsky <joel.rozowsky@yale.edu> Tue, Nov 24, 2015 at 8:16 AM
To: Declan Clarke <declan.clarke@yale.edu>
Cc: Anurag Sethi <anurag.sethi@yale.edu>

Begin forwarded message:

From: "Structure Editorial Office" <em@editorialmanager.com>
Date: November 23, 2015 at 9:08:32 PM EST
To: "Mark Gerstein" <pi@gersteinlab.org>
Subject: Structure Decision

Mark Gerstein, Ph.D.
Yale University
260/266 Whitney Avenue PO Box 208114
New Haven, CT 06520
UNITED STATES

Email: pi@gersteinlab.org

Nov 23, 2015

Manuscript: STRUCTURE-D-15-00408
Title: Identifying allosteric hotspots with dynamics: application to conservation in deep sequencing

Dear Dr. Gerstein,

Thank you for submitting this paper to Structure. We have now received comments on it from the referees. The reports contain a number of comments that would need to be satisfactorily addressed before we could consider the paper further. We would therefore like to invite you to address these comments in a revised version of the manuscript. One of the referees considers the revision to be of sufficient importance to warrant a second review. Accordingly, we will send the revised manuscript back for re-review.

While you are revising your manuscript, please attend to the following editorial points to help expedite the publication of your manuscript.

- **Abstract:** Please make sure that the abstract does not exceed 150 words.

- **Introduction:** Please include a brief summary of your key findings in the final paragraph of the Introduction section.

- **Overall length:** Papers should not exceed 55,000 characters, including spaces. If you need to exceed this guideline please get in touch with us so that we can discuss options.

- **Figures and Tables:** Structure allows a maximum of 10 display items (figures and tables), so please do not exceed this number. For papers describing new structures to atomic resolution, one figure should show the whole molecule or a set of fragments in sufficient detail so that the locations of individual residues can be identified. The Table containing X-Ray data collection, phasing, and refinement

statistics, or the Table containing NMR and refinement statistics MUST be included in the main text.

- **IMPORTANT: Relationship between main text and supplemental display items** - all figure and table legends in the main text must include the information about what supplementary information display items they are supported by. The same applies for the supplementary figures' and tables' legends that must specify which main text item they are supporting.
For example, the last sentence in the main text Figure legend should be: "See also Figure S1.", while Figure S1 in return should be named: "Figure S1, related to Figure 1."
Please check more detailed information on how to structure Supplemental Information here:
http://www.cell.com/supplemental-information

- **Data deposition:** Atomic coordinates, EM data, NMR assignments, nucleic acid sequences, and protein sequences should be deposited in an appropriate databank, and the entry code or accession number must be cited.

- **References:** All in-text citations and listed references should be set in Cell Press style: in-text citations should include first author name and year of publication (Smith et al., 2003), and the reference list should appear in alphabetical order. For examples of the new format, please visit http://www.cell.com/structure/authors#prep.

- **Highlights:** Please include up to 4 highlights, up to 85 characters each.

- **Graphical Abstracts:** Although optional, we strongly encourage the authors to include them.

Please send your revised manuscript via EM as quickly as possible with a cover letter stating clearly how the text has been changed in response to the comments or giving your reasons for not making some of the suggested changes. If your paper is accepted for publication, you will be asked to upload a final submission via EM.

Please do not hesitate to contact us if you have any problems or questions.

Kind Regards,

Andrej Sali, Editor
sali@salilab.org

On behalf of the editors
Bob Fines
Interim Journal Associate, *Structure*
50 Hampshire St., 5th Floor
Cambridge, MA 02139, USA
617-386-2128
structure@cell.com

Reviewer #1: This manuscript presents what seems to be a useful method. Even though the authors highlight deep sequencing, in practice it is a 3-D method. To predict allostery/allosteric residues one needs structures.

The approach itself is not novel. It is a modified version of an earlier one (by Berezovsky et al), with the modifications appearing to efficiently filter and trim the output. Modeling the protein as a network, with residues representing nodes and edges representing contacts between residues is not new either, and

neither is the analysis of residue conservation in the networks. The finding that allosteric residues are significantly conserved over both long and short evolutionary time scales is also not new and indeed expected, as is the observation that not all conserved residues can be explained by protein-protein interactions or in close-packed hydrophobic core.

Despite this lack of conceptual novelty, the usefulness of the paper whose main thrust is the efficient streamlined method, its broad application and its availability can merit its publication. Allostery and allosteric residues and their identification is gaining increasing interest in the community. Having the atlas that they produced along with an efficient accessible method is important.

I have only a couple of minor comments. With regard to conserved residues, networks, information and communication, it would be appropriate to cite an early paper in this direction, Mol Syst Biol. 2006;2:2006.0019. Residues crucial for maintaining short paths in network communication mediate signaling in proteins.(PMID: 16738564). Additionally, though a different implementation, still the papers by S. Vishveshwara (e.g. Biochemistry. 2008 Nov 4;47(44):11398-407. doi: 10.1021/bi8007559) also deserve citing. I would also suggest to the authors to reconsider their title. Even though I understand their wish is to highlight "deep sequencing", some readers may find this title confusing, since eventually the authors use structures.

Reviewer #2:
This paper presents the reimplimentation of two methods for detecing allosteric sites, a server/program for applying the method, the application of the methods to a number of proteins, and an evaluation of the conservation of the identified residues. There are a number of issues with the manuscript which I would like to see addressed.

Major issues:
How were the 12 'canonical' systems chosen? A quick check of a couple of them indicated to me that the functional role of the ligands in allostery has been established. If this is the case for all of them, I think it would be of benefit to the reader to indicate this.

In the supplementary methods for the MC search, although an attractive potential in the -0.05 to -0.75 range is sampled, it is unclear what the repulsive and strongly repulive energies were. The same as the Mitternacht and Berezovsky values (3 and 10)? These are not stated, but would have a significant effect on the sampling. I am also a little confused as to what else is being optimised in the MC scheme. As far as I can tell it is just one parameter, the depth of the well, but the text refers to an "optimal set of parameters" and a "combination of parameters" which best identifies known ligand binding sites.

There appear to be a couple of important steps missing from the supplementary methods. For instance, how is the MC ensemble turned into a list of sites? How are the leverage scores for these sites calculated?

It is difficult to gauge the strength of the predictions in Table S2. For instance, for 2hnp, 67% of the residues are predicted as surface-critical, but over 20% of the residues are buried. Although this is the extreme case, it seems odd to include the interior residues when calculating the fraction of predicted residues and the fraction of ligand-binding residues, when these residue are a priori excluded from both lists. I think it would be more meaningful to report the fraction of surface residues predicted within critical sites, the fraction that are known ligand-binding residues, and the overlap between these two sets, as well as the number of critical sites identified, number of binding sites and the number of strongly overlapping sites. This would make table 3 redundant, put all the relevent information in the same place, and greatly aid interpretation.

"... the mean fraction of GN-identified interior-critical residues that match Infomap-identified residues is 0.30 (the expected mean, based on a uniformly-random distribution of critical residues throughout the protein, is 0.21, p-value=0.058), further justifying our decision to focus on GN)" - I am unclear how this adds to the justification for choosing GN over Infomap.

The paper appears unbalanced. An unusually large effort is dedicated to explaining, illustrating and

analysing the structural clustering scheme, including a section in the main text, figure 2C-E, supplementary figures S8, S9, S10, S21, S22, S23, and over three pages of supplementary methods. The purpose of all this, it seems, is to apply the interior and surface critical methods using these motions instead of the ANM modes. However, how this in done is barely described. How is a set of representative cluster members turned into the equivalent of NMA eigenvectors? Both the surface- and interior-critical method use 10 eigenvectors, but it appears that there are always fewer than 10 cluster members for all proteins investigated, with the reader left to spectulate on how this discrepency is resolved. The results of this extended application only appear in the main text as a pointer to supplementary figure S17.

All ConSurf scores are normalised to zero, but is the variation also set to unity?

Minor issues:
There is an asterix next to two entries in Table S2, and next to one entry in Table S3, but these are not explained in the captions or the main text.
"allosteric ligand has a global affect on a protein's functionally important motions" affect -> effect
jaccard -> Jaccard, three occurrances
line 279: "However 1000 Genomes SNVs tend hit..." -> tend to