RESPONSE LETTER

We thank the reviewers for carefully reading through our study, as well as for valuable feedback on how this work may be improved. Below, we respond to the various issues raised. Before addressing each of these points individually, however, we highlight some of the more global changes that have been introduced to better conform to the format guidelines in *Structure*, as well as to improve readability overall. Such changes include:

- the introduction of more main text figures (these figures were originally in the Supplement)
- more details regarding some of the methods and their implementation
- more contextual language and perspective around the methods described within the Supplement
- numbered sub-headings within the Supplement (as well as a more local sub-headings scheme), thereby making it easier to find information and reference other parts of the Supplement; the overall layout is given in the first page of the Supporting Information

Reviewer #1

-- Ref 1.0 – Emphasis on Deep Sequencing --

Reviewer Comment	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 … 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.
Author Response	We agree that the method is fundamentally 3-D structural in nature, and we feel that readers would have the same reaction as the reviewer. Thus, we have changed our title accordingly. The revised title is "Identifying allosteric hotspots with dynamics: application to inter- and intra-species conservation"

-- Ref 1.1 – General Comments Regarding Novelty and Value --

Reviewer Comment	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.
Author Response	We thank the reviewer for these comments. It is true that the allosteric prediction methods themselves are not fundamentally novel. It is our expectation, however, that our streamlined pipeline and publically available server and source code will facilitate the identification of allosteric residues throughout the protein surface and interior. In addition, we anticipate that the atlas provided may further motivate other studies into allosteric residues on a database scale. In addition, we have now done more to highlight our tool and workflow by including its associated images within the main text (now Fig. 3 in the main text).

-- Ref 1.2 – Citing Early Work on Network Analysis --

Reviewer Comment	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.
Author Response	We thank the reviewer for bringing these studies to our attention, and we now introduce these works within the main text. Specifically, we mention the study by Ghosh *et al*. as part of our introduction to previously developed methods, and we discuss some of the key findings of interest by del Sol *et al*. within the discussion.
Excerpt From Revised Manuscript	… Ghosh et al. have taken a novel approach of combining MD and network principles to characterize allosterically important inter-domain communication in methionyl tRNA synthetase (Ghosh et al., 2008). …In one of the early studies employing network analysis, del Sol *et al*. conduct a detailed study of several allosteric protein families (including GPCRs) to demonstrate that residues important for maintaining the integrity of short paths within residue contact networks are essential to enabling signal transmission between distant sites (del Sol *et al.*, 2006). Another notable result in the same work is that these key residues (which match experimental results) may become redistributed when the protein undergoes conformational change, thereby changing optimal communication routes as a means of conferring

	different regulatory properties.

Reviewer #2

-- Ref 2.1 – Selection of 12 Canonical Systems --

Reviewer Comment	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.
Author Response	Given the importance of the canonical set in our study, we thank the reviewer bringing this ambiguity to our attention. We have clarified the motivating factors behind our choice of a canonical system, and this clarification is now provided in the caption of Table S1, where we fully list the proteins and their ligands (a pointer to this rationale is also given in the main text).
Excerpt From Revised Manuscript	**Table S1: Set of 12 canonical proteins, organized by state (apo or holo)** These 12 proteins were chosen to constitute the canonical set for several reasons: the allosteric mechanisms of their natural ligands are well understood, and both the holo and apo states for each system are available and clearly distinguishable; in addition, these proteins have been extensively investigated in the contexts of both binding leverage and allostery in general. Ligands are given in parentheses (those in bold text designate the ligands used to define residues involved in ligand-binding interactions).

-- Ref 2.2-1 – Parameterization Values --

Reviewer Comment	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 repulsive 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.
Author Response	We thank the reviewer for bringing it to our attention that these details were missing, as the parameters and the means of optimizing them are essential to how surface-critical residues are identified. We have now clarified these items in Supplementary Methods section 3.1-a-i.
Excerpt From Revised Manuscript	… the optimized set of parameters were as follows (here, $D_{lig\text{-}prot}$ designates the distance, in Angstroms, between a ligand atom and a protein atom): *widths* — *depths & heights* $\infty > D_{lig\text{-}prot} \geq 4.5$: Energy = 0 $4.5 > D_{lig\text{-}prot} \geq 3.5$: Energy = - 0.35 (attractive) $3.5 > D_{lig\text{-}prot} \geq 3.0$: Energy = +10 (repulsive) $3.0 > D_{lig\text{-}prot} \geq 0.0$: Energy = +10000 (strongly repulsive: effectively prohibited)

-- Ref 2.2-2 – Parameters being optimized --

Reviewer Comment	I am also a little confused as to what else is being optimized 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.
Author Response	We thank the reviewer for pointing out that this was not clear. Again, this is an essential aspect in our search for surface-critical residues. We have now clarified these items in what is now Supplementary Section 3.1-a-i.
Excerpt From Revised Manuscript	Specifically, the parameters to be optimized include (1) the ranges of favorable and unfavorable interactions (i.e., potential function *widths*) and (2) the attractive and repulsive energies themselves (i.e., potential function *depths* and *heights*)… In addition to optimizing these parameters within the potential function, we also determined that setting the number of MC steps to 10,000 times the size of the simulation box (see above) provided the best convergence across multiple simulations on the same protein – that is, this number of steps better enabled us to recapture the same set of sites when running the simulations multiple times.

-- Ref 2.3-1 – List of Sites from MC --

Reviewer Comment	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?
Author Response	We thank the reviewer for bringing this to our attention as well. This information is now provided in 3.1-a.
Excerpt From Revised Manuscript	After all candidate sites are identified by these MC simulations, pairs of sites with extremely high overlap are merged by combining any pair of sites that have a Jaccard similarity of at least 0.7, where the Jaccard similarity between sites i and j is $\|i \cap j\|/\|i \cup j\|$. After merging sites in this way, the residues of a given site are listed by their local closeness, and no more than 10 residues for a site are used. Local closeness (LC) is a geometric quantity that provides a measure of the degree of a residue in the residue-residue contact network; see (Mitternacht and Berezovsky, 2011b) for further discussion of LC. This entire process results in a list of sites on which binding leverage calculations can be performed.

-- Ref 2.3-2 – Calculating Binding Leverage Scores --

Reviewer Comment	How are the leverage scores for these sites calculated?
Author Response	We thank the reviewer for emphasizing the importance of providing this information. Although our previous Supplement points readers to the work of Mitternacht and Berezovsky for the full formulas, these formulas should indeed be provided explicitly in the Supplement. This information is now provided in 3.1-a-ii.
Excerpt From Revised Manuscript	Specifically, the binding leverage score for a given site is calculated as $$\textit{binding leverage} = \sum_{m=1}^{10}\left(\sum_i \sum_j \Delta d^2_{ij(m)}\right)$$ Here, the outer sum is taken over the 10 modes, and the pair of inner sums are taken

	over all pairs of residues (*i*,*j*) such that the line connecting the pair lies within 3.0 Angstroms of any atom within the simulated ligand. The value $\Delta d_{ij(m)}$ for each residue pair (*i*,*j*) represents the change in the distance between residues *i* and *j* when this distance is calculated using mode *m*. Thus, one may think of binding leverage as qualitatively predicting the extent to which a surface pocket is deformed when the protein undergoes conformational transitions… … when using ACT vectors, the binding leverage score for a given site is simply calculated as: $$binding\ leverage\ = \sum_i \sum_j \Delta d_{ij}^2$$ where the sum is taken over all pairs of residues (*i*,*j*) such that the line connecting the pair lies within 3.0 Angstroms of any atom within the simulated ligand, and the value Δd_{ij} for each residue pair (*i*,*j*) represents the change in the distance between residues *i* and *j* when this distance is calculated in alternative crystal structure. Thus, for each residue, the 10 vectors provided by the normal modes are simply replaced by the single ACT vector that defines the change in position of that residue when going from the protein conformation given by one representative structure to the conformation given by the other representative.

-- Ref 2.4 – Table with Statistics on Surface Residues --

Reviewer Comment	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 residues 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 relevant information in the same place, and greatly aid interpretation.
Author Response	We thank to reviewer for raising these important points. We agree that only the surface residues should be included in these calculations, our presentation of this information can be clarified by keeping all of the information within one table, and more statistics would aid in interpretation. Along these lines, we have done the following: • Our analysis has been revised to consider the surface residues specifically (specifically, we define surface residues by using NACCESS [[cite]] to select those residues with a relative solvent accessibility exceeding 50%). • These two tables have been merged and expanded, and additional data (such as the number of known ligand-binding sites) is now included. • We have also decided to move this merged Table from the Supplement into the main text of the manuscript.

Excerpt From Revised Manuscript	Table 1. Statistics on the surfaces of apo structures within the canonical set of proteins

PDB ID	% Surf (SC res)	% Surf (LB res)	SC-LB overlap	# SC sites	# LB sites	# Overlapping sites	% LB sites identified
3pfk	51.0	20.4	0.255 (0.155)	19	3	3	100.0
4ake	45.4	17.8	0.274 (0.154)	29	2	2	100.0
1cd5	58.9	10.0	0.153 (0.096)	24	2	1	50.0
1j3h	6.6	8.0	0.25 (0.041)	2	1	1	100.0
1bks	34.3	9.7	0.079 (0.079)	24	4	1	25.0
1e5x	20.7	9.3	0.139 (0.077)	17	3	2	66.7
1efk	5.5	8.6	0.03 (0.036)	10	10	0	0.0
1nr7	14.9	17.5	0.187 (0.102)	45	24	6	25.0
1xtt	29.8	19.6	0.295 (0.154)	31	5	5	100.0
2hnp	73.9	13.3	0.16 (0.134)	25	2	2	100.0
3d7s	26.7	13.7	0.054 (0.064)	26	9	0	0.0
3ju5	1.6	3.9	0 (0.013)	1	2	0	0.0
mean	30.8	12.7	0.156 (0.092)	21.083	5.583	1.917	55.6

Table 1: Statistics on the surfaces of *apo* structures within the canonical set of proteins

For each *apo* structure within the canonical set of proteins, statistics relating surface-critical sites to known ligand-binding sites are reported. The surface of a given structure is defined to be the set of all residues that have a relative solvent accessibility of at least 50%, where relative solvent accessibility is evaluated using all heavy atoms in both the main-chain and side-chain of a given residue. Mean values are given in the bottom row. *Column 1*: PDB IDs for each structure; *Column 2*: among these surface residues, the fraction that constitute surface-critical residues; *Column 3*: among surface residues, the fraction that constitute known ligand-binding residues (known ligand-binding residues are taken to be those within 4.5 Angstroms of the ligand in the *holo* structure; Table S1); *Column 4*: the Jaccard similarity between the sets of residues represented in columns 2 and 3 (i.e., surface-critical and known-ligand binding residues), where values given in parentheses represent the expected Jaccard similarity, given a null model in which surface-critical and ligand-binding residues are randomly distributed throughout the surface (for each structure, 10,000 simulations are performed to produce random distributions, and the expected values reported here constitute the mean Jaccard similarity among the 10,000 simulations for each structure); *Column 5*: the number of distinct surface-critical sites identified in each structure; *Column 6*: the number of known ligand-binding sites in each structure; *Column 7*: the number of known ligand-binding sites which are positively identified within the set of surface-critical sites, where a positive match occurs if a majority of the residues in a surface-critical site coincide with the known ligand-binding site; *Column 8*: The fraction of ligand-binding sites captured is simply the ratio of the values in column 7 to those in column 6.

-- Ref 2.5 – GN vs. Infomap for Network Analysis --

Reviewer Comment	"... 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.
Author Response	We thank to reviewer for highlighting this ambiguity. Here, the important issue is the fact that GN is far more selective than Infomap in identifying important network elements (i.e., interior-critical residues), as

	evidenced by the data presented in Table S3 (previously Table S4). Furthermore, not only does GN provide a more selective set of residues, but the network modularity given by GN is somewhat better than that provided by Infomap, which is also given in Table S3. These issues have been clarified in SI Methods section 3.1-b-ii.
Excerpt From Revised Manuscript	Although the critical residues identified by GN do not always correspond to those identified by Infomap, 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). Furthermore, we observe that obvious structural communities are detected when applying both methods: a community generated by GN is often the same as that generated by Infomap, and in other cases, a community generated by GN is often composed of sub-communities generated by Infomap. In addition, the modularity from the network partitions generated by GN and Infomap are comparable. For the 12 canonical systems, the mean modularity for GN and Infomap is 0.73 and 0.68, respectively. Presumably, GN modularity values are consistently at least as high as those in Infomap because GN explicitly optimizes modularity in partitioning the network, whereas Infomap does not. Together, these results suggest that both GN and Infomap generate similar partitions. Roughly, the set of interior-critical residues identified by GN partially constitute a subset of those identified with Infomap. If these sets of residues were completely different, then the choice between GN and Infomap would be difficult, as the results in our downstream conservation analyses would then be highly sensitive to our community detection method of choice. Given that the two residue sets are not disjoint, our choice of GN over infomap was largely guided by the fact that GN is far more selective in identifying important network elements (i.e., interior-critical residues), as evidenced in Table S3. In contrast, Infomap generates a much less selective set of interior-critical residues.

-- Ref 2.6-1 – Overemphasis on structural clustering scheme --

Reviewer Comment	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…
Author Response	We thank the reviewer for this observation, and we agree that we had devoted a large amount of our discussion to what is more of a preliminary protocol. Accordingly, we have tried to de-emphasize some of the content related to the structural clustering. Specifically, we have: • moved Fig. 2C-E out of the main text and into the SI (now Fig. S3) • merged what was previously SI Figs. S8, S9, and S12 (along with what was previously Fig. 2C-E) into one SI item (now Fig. S3) • condensed the relevant text in the Supplement (now SI Methods sections 3.2-a and 3.2-b) from 3 pages to less than 2; and • completely removed Figs S10, S21, S22, and S23, which we consider to be somewhat extraneous. We note, however, that because the structural clustering scheme is not

	a previously established method, considerable care had to be devoted to ensuring that it was working as intended. Our discussion regarding the clustering scheme and its importance in this study might be clarified in our response to Comment 2-6.2 below.

-- Ref 2.6-2 – Clarifications Regarding ANMs & ACT Vectors --

Reviewer Comment	…The purpose of all this [structural clustering scheme], 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 speculate on how this discrepancy is resolved. The results of this extended application only appear in the main text as a pointer to supplementary figure S17.
Author Response	We thank to reviewer for highlighting these ambiguities. In our response here, we try to clarify these protocols by first providing the motivating factors behind the clustering scheme. In addition, within the box below, we highlight the text that we have added in order to clarify the implementation of these methods. The purpose of developing and implementing the clustering scheme is three-fold: 1) We are primarily interested in those structures that exhibit distinct conformations, as we are focusing on cases for which pronounced global conformational change play essential roles in allosteric mechanisms. 2) The clustering scheme ultimately enables us to perform an important control. Namely, it enables us to address the question: are the results robust to alternative methods of inferring information about conformational change? ANMs provide only one means of defining the vectors for modeled conformational change. However, another approach is to use the displacement vectors from the crystal structures of alternative conformations. This alternative constitutes a method that we term “absolute conformational change” (ACT). 3) Because ANMs constitute the bulk of our analysis (see below), we must be confident that the structures being analyzed with ANMs are suitable for normal modes analysis: if a given protein is not believed to undergo significant conformational change, it may not be appropriate to apply ANMs, as the ANMs can incorrectly predict large-scale conformational change where no such change is likely to occur.
Excerpt From Revised Manuscript	Unless otherwise specified, we use normal modes analysis to model conformational change throughout this study. However, one potential concern with this approach is that normal modes may not faithfully represent plausible conformational changes. Thus, in order to determine whether or not the results are

	robust to different means of inferring motions (especially those results relevant to the conservation of critical residues), we also model conformational change using vectors connecting pairs of corresponding residues in crystal structures of alternative conformations. We term this approach "absolute conformational transitioning" (ACT). This more direct model of conformational change is especially straightforward to apply to single-chain proteins (applying ACT on a database scale to multi-chain complexes would introduce confounding factors related to chain-chain correspondence between such complexes when each complex has multiple copies of a given chain). **3.2-c-i Inferring Protein Conformational Change Using Displacement Vectors from Alternative Conformations** Given a particular protein, how are these ACT vectors defined in order to calculate critical residues? For demonstration, we discuss a hypothetical example consisting of a multiple structure alignment of 8 sequence-identical structures. Starting with the protein's multiple-structure alignment using all 8 structures, we determine the optimal number of clusters represented by the structure alignment using the K-means algorithm with the gap statistic (see the above SI Methods section 3.2-b). Suppose that these 8 structures may be grouped into 2 distinct clusters by our scheme (4 structures in *cluster A*, and 4 structures in *cluster B*, for instance). As discussed in SI Methods section 3.2-b, a representative structure is taken from each of these two clusters (*structure A* and *structure B*). These two representatives are taken to represent the alternative conformations for the protein. As an alternative to using ANMs, we may use *structure A* and *structure B* to try to infer information about the protein's global conformational shifts by assigning a displacement vector to each residue (for instance, residue Y140), where the displacement vector is simply defined by the two corresponding residues in the different structures within the structure alignment (i.e., Y140 within *structure A* of the structure alignment and Y140 within *structure B* of the structure alignment). Because the structure alignment was performed on sequence-identical structures, each residue in one of these two representative structures matches a corresponding residue on the other representative structure. If each of the two structures represents a sequence-identical protein consisting of 200 residues, then 200 ACT vectors are drawn in order to represent the conformational change in transitioning from one conformation to the other. These 200 ACT vectors for the protein may then be used to identify surface- and interior-critical residues (see below), and downstream analysis on these residues is then performed.

-- Ref 2.7 – ConSurf Normalization --

Reviewer Comment	All ConSurf scores are normalised to zero, but is the variation also set to unity?
Author Response	We thank to reviewer for noting this omission. Indeed, value for σ^2 is set to unity, and this is now indicated in Supplement Section 3.3-a.
Excerpt From Revised Manuscript	ConSurf scores for each protein chain are normalized to have a mean ConSurf score of 0 (the ConSurf score variance is 1 for each chain).

-- Ref 2.8 – Minor Issues --

Reviewer Comment	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
Author Response	We thank the reviewer for pointing out these points. With respect to the asterix symbols in Table S2, and next to one entry in Table S3 (now merged into what is now Table 1, as noted), these were originally intended to highlight structures for which the identification of biological ligand-binding sites was previously known to be especially difficult. However, this information is not essential, and may be distracting. Thus, the asterix symbols have been removed, and this is no longer considered. We have also corrected the other two issues raised here, and thank the reviewer again for a very careful review of this work.