Following text explains the choice of parameters and their impact on the analysis of conserved clusters in CCRXP.

1. **Choice of residue position assignment in space:** For the purpose of clustering conserved residues in a protein, each residue is replaced by a point, whose position can be determined by either using its Cα-atomic position or by taking the geometric center of all its constituent atoms. In general, the two choices produce very similar results. Option to choose any one of them has been provided to deal with unusual situations that might arise. For example, if there are too many missing atoms in a protein, geometric center will not be an appropriate choice as it will assign inconsistent positions to residues. On the other hand Cα atomic positions ignore side chains altogether, and some residues connected by side chains contacts may be missed. So, if one is sure about the structural quality of his structures and there are no missing atoms in the submitted file, geometric center may produce more robust results. Default is Cα-atoms.

2. **For calculating packing density:** Packing density is calculated by counting the number of residues within a given distance from each query residue. This requires assigning each residue to a point in space just as we did for clustering above. Since, packing density is NOT a clustering parameter; this choice will not affect clustering at all. However, packing density values returned as additional features of clusters/ residues by the server may be slightly affected, based on the same considerations as above. Default is geometric center.

3. **Distance criterion, for packing density calculations:** As stated above, packing density is calculated by counting the number of residues within a given distance from each query residue. Average packing density of a cluster is calculated by taking the mean of packing density at the site of each residue in the cluster. In our published work, where we showed that the hot spot residues are tightly packed, we have used a distance cutoff=7Å, which is the server default. If, the distance is less than this, we may get too few residues to get a reliable packing density score, whereas, at larger distances, packing density of all query residues may become the same. Unless, users are interested in looking at specific packing patterns in the proteins, the default value is fine. However, this value does not affect which

residues will belong to a cluster, as this is just an additional feature of the cluster and not used for clustering itself.

4. **Distance criterion, for cluster joining:** This is the most significant parameter in the server and affects the residue populations and all aspects of clusters. When clustering is performed a residue is assigned to a cluster if any residue in the current cluster is at a distance less than this cutoff (single linkage method). We have analyzed clusters in our previous analyses at 5Å cutoff criterion and that's the default value used. Users should change this only if they have clear reason for doing so. For example, in some cases there may be too many residues in a custer at this cutoff and clusters may be better analyzed by choosing a more strict distance cutoff (smaller value of cluster joining distance). This may be true for proteins, which have too many conserved residues. On the other hand joining distance may be relaxed if there are very few conserved residues in the structure and long-range interactions are to be investigated.

5. **Definition of residue contact with DNA:** This parameter is applicable only to DNA-binding proteins to analyze the number of DN-contacts of each residue or cluster. In all our studies, we have used 3.5 Å any atom to any atom distance criterion to define DNA-contacts and that's the default. However, different researchers have used other cutoffs and that's why a choice has been provided. This parameter does not affect the members or size of the clusters.

6. **Conservation score options:** Final clusters are sensitive to this score. We used a third-party software *Scorcons* by Valdar as cited in the main text to calculate conservation scores. The score returned by this program assigns a value between 0 and 1 for each residue (0 for no conservation and 1 for 100% conservation). Choosing a large cutoff leads to small clusters of highly conserved residues and choosing a small cutoff will make larger clusters with less strong conservation. In our analyses we have used 0.7 as cutoff score to analyse DNA-binding residues and 0.6 for analyzing protein-protein interfaces. Scores were relaxed for protein-protein interactions to have statically significant data in conserved and non-conserved categories. Default value of 0.7 may be enough for most analysis, but depending

upon the overall conservation of a protein sequences within the family, users may need to make a more strict or relaxed choice, depending upon the purpose of their analysis.

7. **Maximum number of aligned sequences**: In order to calculate conservation scores, a database search is performed to find sequences similar to the query protein. In some cases sequence database returns too many similar sequences and multiple alignments required for calculating conservation scores become too slow if too many sequences are included. In that case, it is necessary to restrict the number of sequences to a smaller number. However, if the top-aligned sequences are very similar to each other, all residues in alignment may be marked as conserved. This can be corrected by including more sequences in the multiple alignments. In most cases, default of 50 sequences works fine, but if the users notice too many conserved residues, this parameter may be tweaked.

8. **ASA cutoff:** This option is provided to identify clusters on surface. If a cutoff other than 0 is selected, only the residues with ASA> cutoff will be considered for clustering. This parameter is provided only for special needs that may arise and by default, all residues are included.

9. **Minimum number of residues in a cluster:** This is just a display option. If there are too many clusters, one may want to focus on only the larger ones. One may also want to remove single-member clusters. If a cutoff is selected, clusters with less number of residues will not be shown in the final output.