CGAP

The visualization toolkit, used for the analysis of statistical residue contact potentials and orientation dependencies can be started (within the drop down menu) via right click on a residue contact within the contact map. The contact map is used to select the residue-residue contact (see Figure 1), which is under investigation, and simultaneously to extract the neighborhood describing String (nbhString).

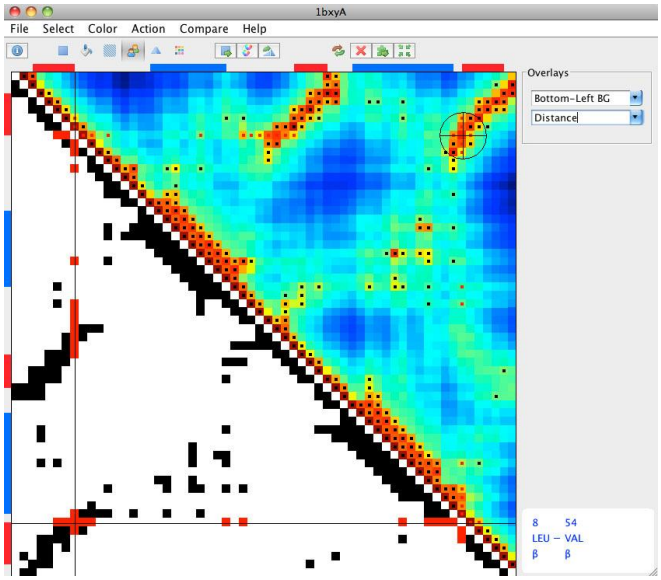


Figure 1: CMView: neighborhood selection of contact within contact map (bottom left) and distance map (top right) of a model for the protein 1bxy, protein chain A.

Sphoxel-Map Representation

As indicator for the residue contact probabilities at specific positions around the central residue, so-called log-odds-scores have been computed. The positions are defined by the two dihedral angles λ (lambda) and ϕ(phi) and the radii restricting the shell, which can be either close, middle or far shell. The user can switch between those shells with the help of a slider. Based on the mapping of the computed LOSs onto a spherical surface, different map projections have been implemented:

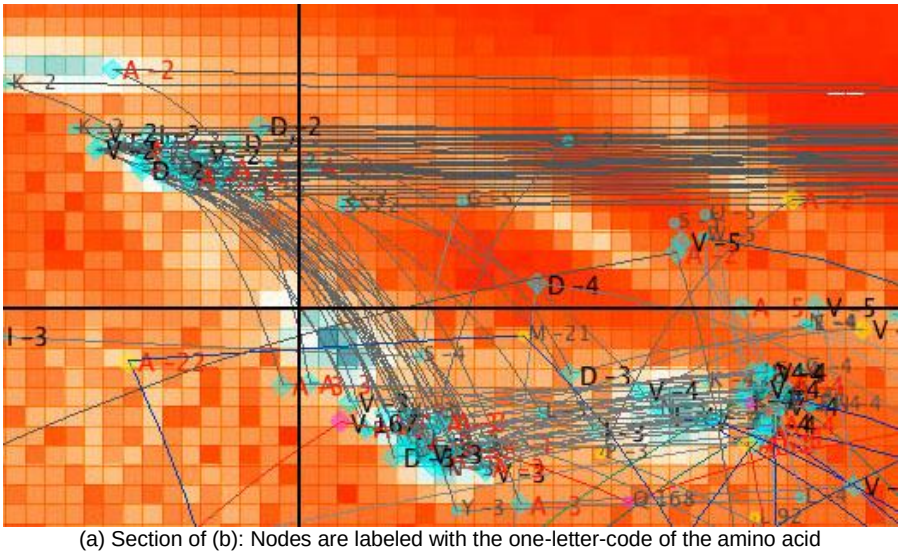
- equidistant cylindrical Plateé Carreé projection
- pseudo-cylindrical KavrayskiyVII projection
- orthographic azimuthal projection

The user can switch between these projections interactively, at the same time changing the rulers, showing exact angle values (see Figures 2). For an easy orientation rulers and a cross-hair, as well as longitudes and latitudes have been implemented. Further, all exact information about the selected contact (about residues I and J) as well as the mouse position (coordinate [λ,ϕ]) are shown within a small information panel at the bottom of the menu (see right bottom of the interface in Figure 6). The negative and positive scores are mapped onto color and the tiles are colored respectively. Three color scales have been implemented:

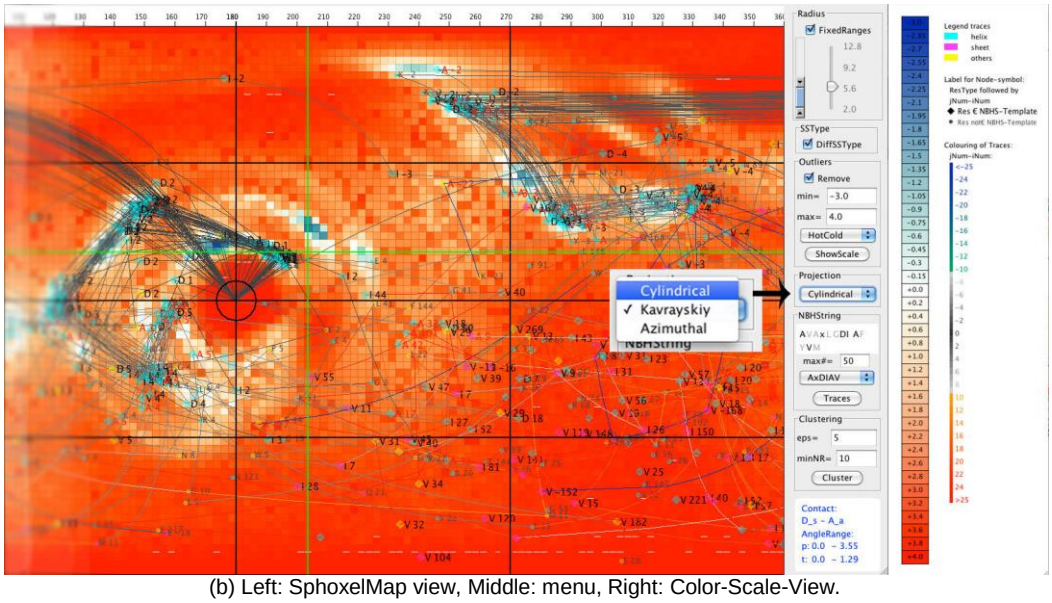
- Blue-Red Scale
- RGB Scale
- Hot-Cold Scale

The scale can be selected by the user via a drop-down box within the menu. The default blue-red scale maps negative scores onto blue, whereas the brightness of the color is determined by its absolute value. Positive scores are thereby mapped onto redwith an brightness depending on the value. When using the RGB scale, negative values are mapped

onto colors from blue to green and positive scores from green to red. In comparison, the hot-cold scale maps negative values onto cold colors only, from blue to green, and positive values onto warm colors, from yellow to red (see e.g. Figure 2). A color-scale-view can be opened with the help of the menu. This view shows how the LOS values are mapped onto color (see Figure 2) and automatically adjusts, when thresholds or the color-scale are changed. It further contains a legend, explaining the shape, color and labels of nodes as well as the coloring of the edges.
The scores are first normalized to a fixed scale [-1:1] before they are mapped onto color. Furthermore, cutoffs can be defined by the user via the interface, which is not only useful to remove outliers, but also to use consistent scaling (normalization) and make different maps for different contact types comparable. Figures 2 and 5 indicate, how the representation changes, when cutoffs are applied. The values can be mapped threshold-based or non-threshold-based. The Scale-View shows how the values are mapped onto color (see Figure 2 or 3).



(a) Section of (b): Nodes are labeled with the one-letter-code of the amino acid type and sequence distance. Bold, black labels for residues of the NBHString, red labels for residues of same type as jRes and smaller grey labels for all others.



(b) Left: SphoxelMap view, Middle: menu, Right: Color-Scale-View.

Figure 2: Cylindrical projection of contact potentials between Aspartic Acid and Alanine with a scaled hot-cold color-scale, i.e. thresholds (-3,+4) were applied. The SphoxelMap representation is superimposed with traces for the template String Ax-DIAV, including the position of the central residue.

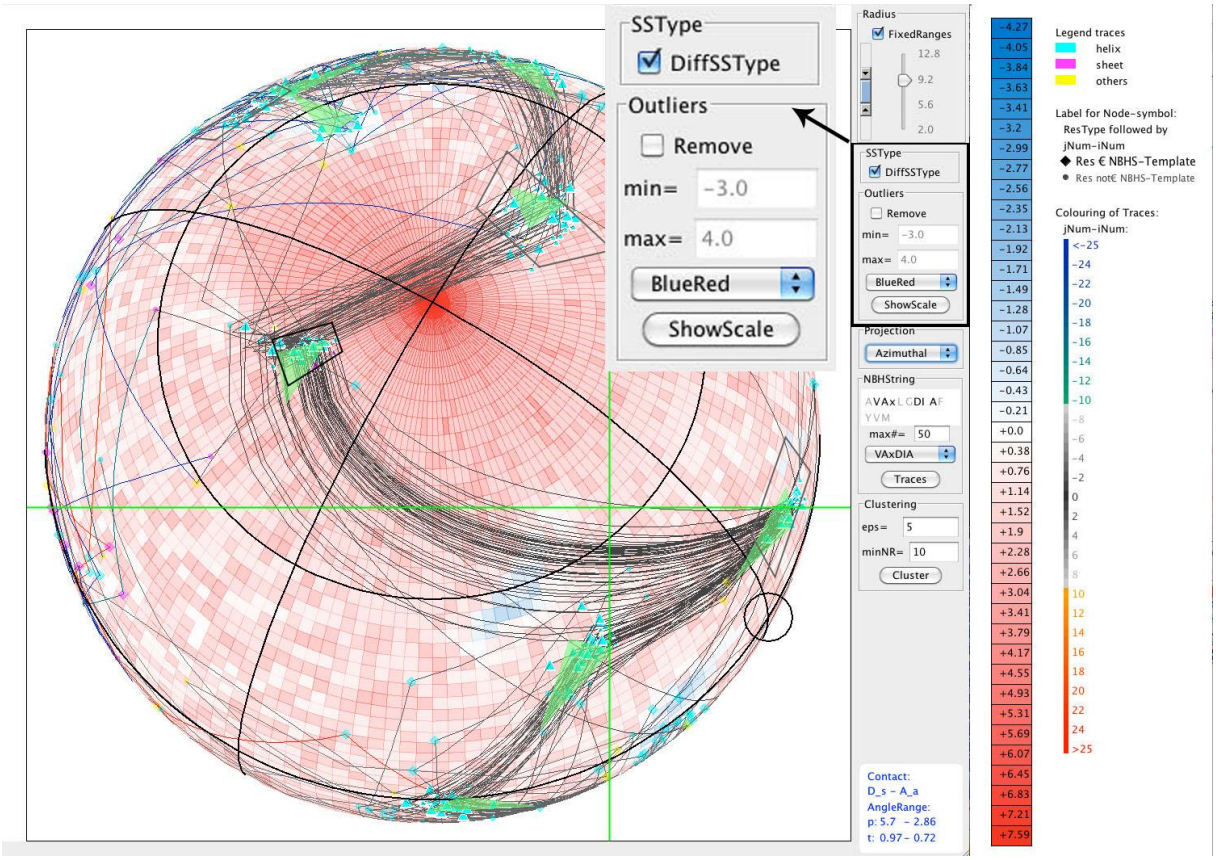


Figure 3: Azimuthal projection of contact potentials between Aspartic Acid and Alanine with a blue-red color-scale. The Neighborhood-Traces are plotted for the template String VAxDIA, excluding the position of the central residue. DBScan was applied and the average direction of outgoing edges from clusters is represented through green arrows. Already selected dihedral angle ranges are surrounded by a grey or black(for the currently selected contact) rectangle. Right: Color-Scale-View.

Neighborhood-Traces

Derivation of template NBHStrings As the original template neighborhood string (nbhString) usually does not provide lots of output traces, it is necessary to reduce the nbhString and remove some residues to increase the number of output traces. It is therefore possible to toggle on or of certain residues interactively (via clicking on the letters within the NBHString selection panel) or via selecting one of the suggested substrings (see Figure 4). The ten best strings, i.e. with the highest support for iRes are offered within a drop down box, from which they can be selected. Thus, it is possible to select a template string from the drop down box and further change it manually. The shorter the nbhString, the more traces are extracted by the query. Nevertheless, to stem the amount of traces for a well scoring NBH template, a maximal number of traces is extracted and visualized. This number can be adjusted via the interface (see Figure 4). The bundles of NBH traces can further be filtered for the secondary structure type (any ssType, helix, strand, loop or others) via another drop down box to analyze the local environments of only these residues that are of some specific ssType.

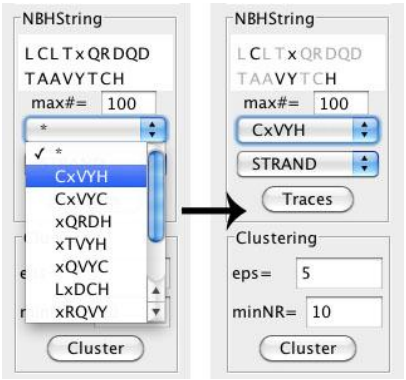


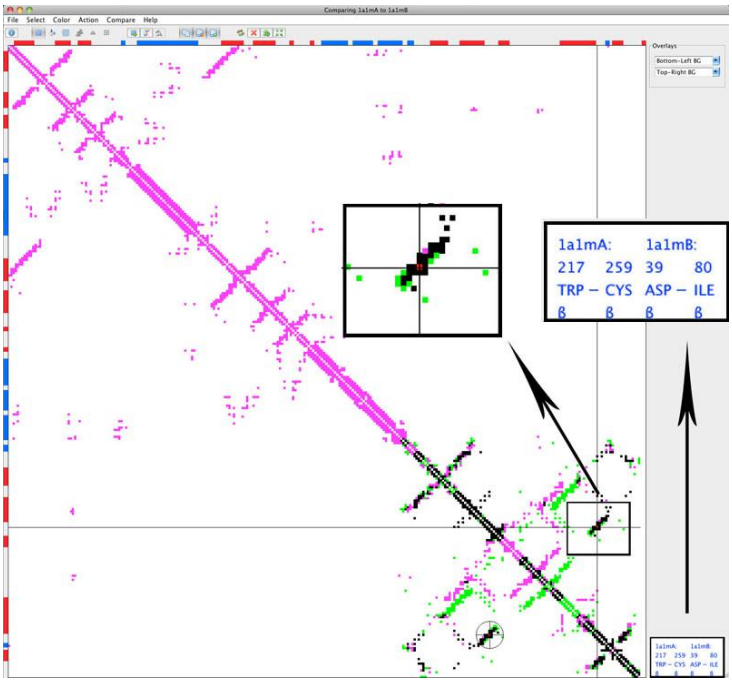
Figure 4: Selection of template-neighborhood-String (nbhString) via drop-down box. Single residues can be toggle on/o_ via clicking on them. Enabled residues of the original (full) template nbhString are colored black and disabled grey.

Nodes of NBH Traces Similar neighborhoods are visualized via plotting the contacting residues at their position with respect to the central residue (see Figure 2). They are connected via edges corresponding their residue number (from lowest to highest number). Thus we get a path (trace) for each extracted neighborhood. The contacting residues are plotted as nodes of different shape, either as sphere (default representation) or as rhombus, if the residue is element of the template nbhString. If the node is element of a cluster (based on the cluster analysis), it is represented as triangle. Not only shape, but also color and size are used for highlighting. Thus, the triangles (nodes of a cluster) and rhombuses (residues of the neighborhood) are bigger. The triangles are further highlighted through a white border. The color is used to represent the secondary structure type of the residue, which might be either helix (cyan), sheet (magenta) or any other SStype (yellow) (see Color-Scale-View in Figures 2 and 3). The nodes are further labeled with the oneletter-code of the residue type and the difference in residue number (jRes-iRes), which represents the distance within the sequence. Thus, all residues that are situated within the sequence before the central residue are labeled with a negative number and those behind with a positive number. These captions can be switched on or off to reduce the amount of represented information.

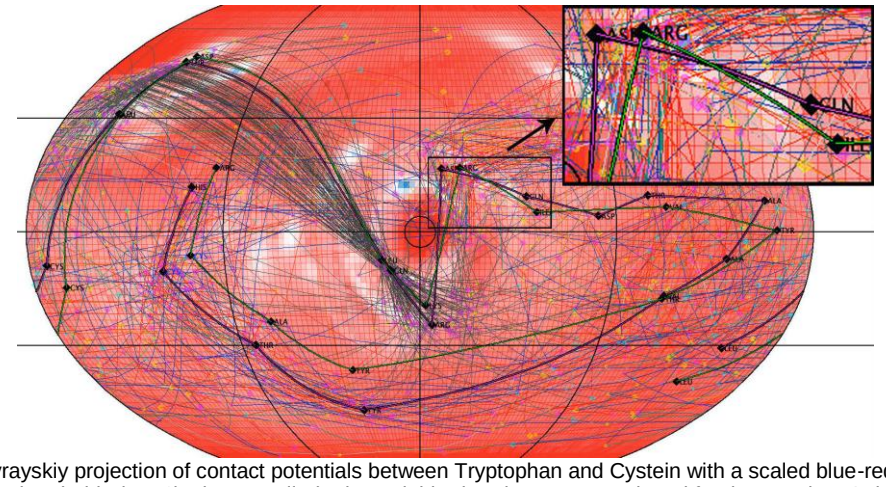
Edges of NBH Traces A color gradient is used for the edges, to visualize where and how far in sequence the residues of the neighborhood are situated. Particularly, the edges between residues, which have a sequence distance smaller than zero, are colored with the cold part of the hot-cold-scale (from blue to green), whereas those with positive distance are colored with the warm colors of the same scale (yellow to red). This colorscheme is just applied onto the long-range contacts, i.e. those contacts that are far in sequence (i.e. |jRes-iRes| > 9A) but close in space. The edges between the short-range contacts are colored based on a grey scale.
The traces can be plotted in two ways either including the central residue or excluding the central residue (see Figure 2 or 5). If the option for the central residue is toggled on, the node is included in the trace and plotted at the centre of the map projection. The edges are thereby implemented and drawn as geodesics instead of straight lines.

Template NBH Traces The extracted neighborhood traces represent statistical information. To analyze how far the local neighborhood of the selected contact complies those statistical spatial propensities, the trace for the template nbhString is highlighted and drawn with bold edges (see Figure 5), if the option is toggled on. In case that two models are loaded in CMView for the purpose of comparison, two traces are plotted within the representation (see Figure 5). This should help to visualize and analyze the local differences in folding of a specific residue neighborhood between the two models of the same protein. The bold traces for the two models are colored appropriate the contacts within the contact map (see Figure 5). As contacts, that are contained only within the first/second model are colored pink/green within the contact map, the related traces are colored respectively. This visualization could be used e.g. in the context of CASP (Critical Assessment of protein Structure Prediction), to

evaluate the accuracy of predicted structures in comparison with the resolved target structures. Thereby, the two models need to be loaded and aligned first. To show the neighborhood traces, the user then needs to select a contact, which both models have in common (one of the black contacts).



(a) Contact map of 1a1m chain A and B, whereas common contacts are black, and those contained only in A pink and in B green.



(b) Kavrayskiy projection of contact potentials between Tryptophan and Cystein with a scaled blue-red color-scale, i.e. thresholds (-4,+6) where applied. The Neighborhood-Traces are plotted for the template String CxVY, excluding the position of the central residue and the template Traces itself (bold and pink/green for chain A/B). Traces are filtered for sheet-type.

Figure 5: Possibilities within the comparing mode of CMView.

Interaction

Change of Views The representation of the LOSs (Sphoxel-Map) as well as of the neighborhood traces can be changed via interactions, particularly via filtering (changing of query parameters, as explained before). The map view can also be changed via panning, i.e. the area of interest can be moved into the centre of the screen with the help of the mouse or the arrow-keys. The navigation within the cylindrical and pseudo-cylindrical map is therefore limited to one dimension, allowing to rotate the sphere around the azimuthal (vertical) axis. In contrast, when using the azimuthal map projection, rotation can be performed around the centre of the sphere (see Figure 3).

Derivation of Orientation Constraints Besides filtering and navigation, interaction is also necessary to define certain orientation constraints. Thereby, it is possible to select a range with the help of the rectangle or cluster selection tool. The latter will be explained within the next paragraph. When using the rectangle selection mode, a range can be defined via clicking on a point and dragging the mouse, while the rectangle becomes visible. During this selection procedure the view is overlaid by a crosshair marking the mouse position and the exact angle position is shown within the information panel (see right bottom of the interface in Figure 6) while the mouse moves over the map. The clicked position is shown on the left and the dragging position on the right of the information panel. This panel also holds information about the currently selected contact, i.e. iResidue and jResidue and their SSType (secondaryStructureTypes). In case, that a restriction range has already been assigned for the currently selected contact, this is replaced by the newest selection. All assigned ranges are visualized within the map via (projected) rectangles (see Figure 3), whereas the rectangle for the currently chosen contact is drawn in different color (black and the rest in grey). A selected range can be deleted via right-click onto the selection and choosing the option DeleteSelectedRange. It is further possible to open a histogram view for a specific range via the drop-down menu (see paragraph Histogram View).
As described before, the Sphoxel-Map as well as the neighborhood representation both depend on the selected contact within the contact map of CMView. Thus, if another contact is clicked within the map, contact constraints and template nbhString change. Thereby, both representations change respectively, as they rely in on these properties, while so far selected ranges are saved for certain contacts. This way it is possible to define various orientation constraints after each other with the help of certain background information depending on the contact.

DBSCAN on Neighborhood-Traces

Clustering Method The DBSCAN algorithm was implemented as clustering technique to extract clusters of nodes of the neighborhood traces. The clustering output can be controlled via the two parameters ϵ (epsilon) and minNumPts (see Interface Clustering, e.g. within Figure 6). ϵ should be declared in degree, preferably within a range from 3°to 8°. The minimal number of points (minNumPts) should have a value within the range of [1 : numberOfTraces].

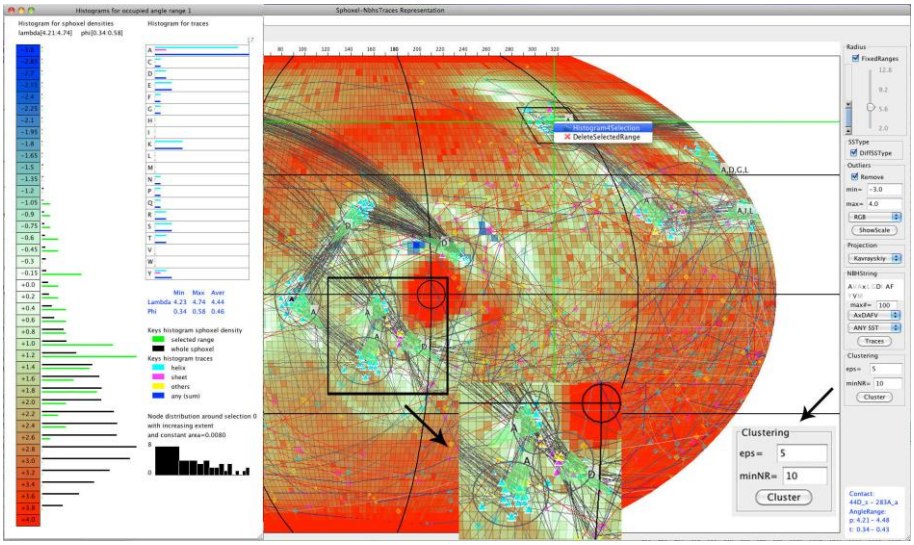


Figure 6: Kavrayskiy projection of contact potentials between Aspartic Acid and Alanine with a scaled RGB color-scale. The Neighborhood-Traces are plotted for the template String AxDAFV, excluding the position of the central residue. DBScan was applied, resolved clusters are highlighted, predominant residue types are plotted on semitransparent white ground and the average direction of outgoing edges from clusters is represented through green arrows. The currently selected dihedral angle ranges is surrounded by a black rectangle. The histogram view for that range is opened via selection within a drop-down menu. Left: Histogram-View for selected angle range.

Visualization of Extracted Clusters The clustering method should help to extract orientation constraints and to determine ranges for the dihedral angles lambda and phi. For this purpose, all nodes that are contained within a cluster are highlighted. As already mentioned in the last paragraph, they are drawn as triangles, surrounded by a white border. To better convey the range of λ and ϕ they comprise, each cluster is surrounded by an ellipse with an extent with respect to the angle ranges (see Figure 6).
Based on this cluster result, a further analysis of the variation in node positions within the cluster is performed, i.e. the minimal, maximal and average values of λ and ϕ are printed within the histogram view (see next paragraph). In addition, the average direction of the outgoing edges is computed for each cluster (with the help of the angle bisectors, i.e. via vector addition and normalization). It is displayed via thick green semitransparent arrows (see Figure 6).
Besides the average direction, also the predominant residue types of each cluster are visualized (see Figure 7). The derived set of clusters allows visual support to determine orientation constraints on the one hand side, but can also be used to automatically derive angle ranges (see section 4 Interaction). For the latter, the user has to switch to the cluster-selection mode. A click on any cluster then automatically sets the angle range for the currently selected contact to the range stretched by the minimal and maximal positions of nodes occurring within the cluster. The rectangle that defines the dihedral angle range will than span around the cluster.

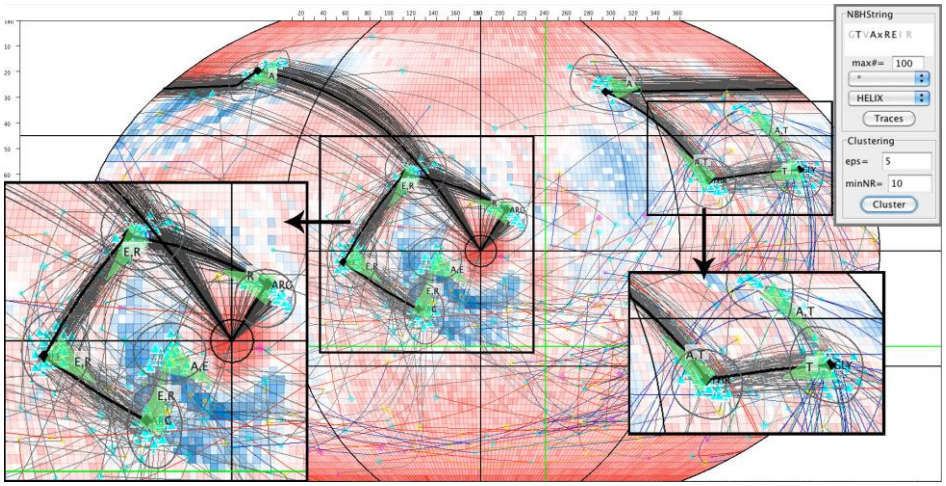


Figure 7: Kavrayskiy projection of contact potentials between Leucine and Glutamic Acid in Blue-Red color-scale. The Neighborhood-Traces are plotted for the template String TAxRE of type Helix, including the position of the central residue. The template NBHString is derived for the contact from protein 1kx5. DBScan was applied, resolved clusters are highlighted, predominant residue types are plotted on semitransparent white ground and the average direction of outgoing edges from clusters is represented through green arrows.

Histogram-View

The histogram-view can be opened for any determined dihedral angle range via right click within a rectangle and choosing the option Histogram4Selection. The histogram-view contains various information and different histograms for the LOSs as well as neighborhood traces, based on the selected range (see Figures 6 and 8).
Histogram of LOSs The histogram for the spatial propensities plots the distribution of LOSs within the selected angle range (green, see legend within histogram-view) in comparison to the whole map (black), i.e. λ[0 : 2π] and ϕ[0 : π]. Thereby, the numbers of occurring pixels, which lie within a certain value range, are determined and plotted via bars next to the color-scale view. The length of the bars is proportional to the ratio of occurring number and maximal number within the selection (for green bars) or whole map (for black bars). For easier interpretation, the bars are not just plotted along a labeled axis for the LOSs, but along

a color-scale that is superimposed with the exact LOS values. The frequency of elements of the histogram helps to determine how favorable this dihedral angle range is.

Histogram of NBH Trace Nodes The second histogram (see right upper histogram of Figures 6 and 8) is based on the neighborhood traces. Within that histogram the numbers of occurring nodes of certain residue type and certain secondary structure type within the selected range are plotted as bars. The bars are colored respective the secondary structure type, similar to the color coding used for the nodes itself. This histogram can be used to determine the predominant residue types within the cluster. Thus it can be used to check whether the selected angle range poses a good constraint for the currently selected contact between iResidue and jResidue.

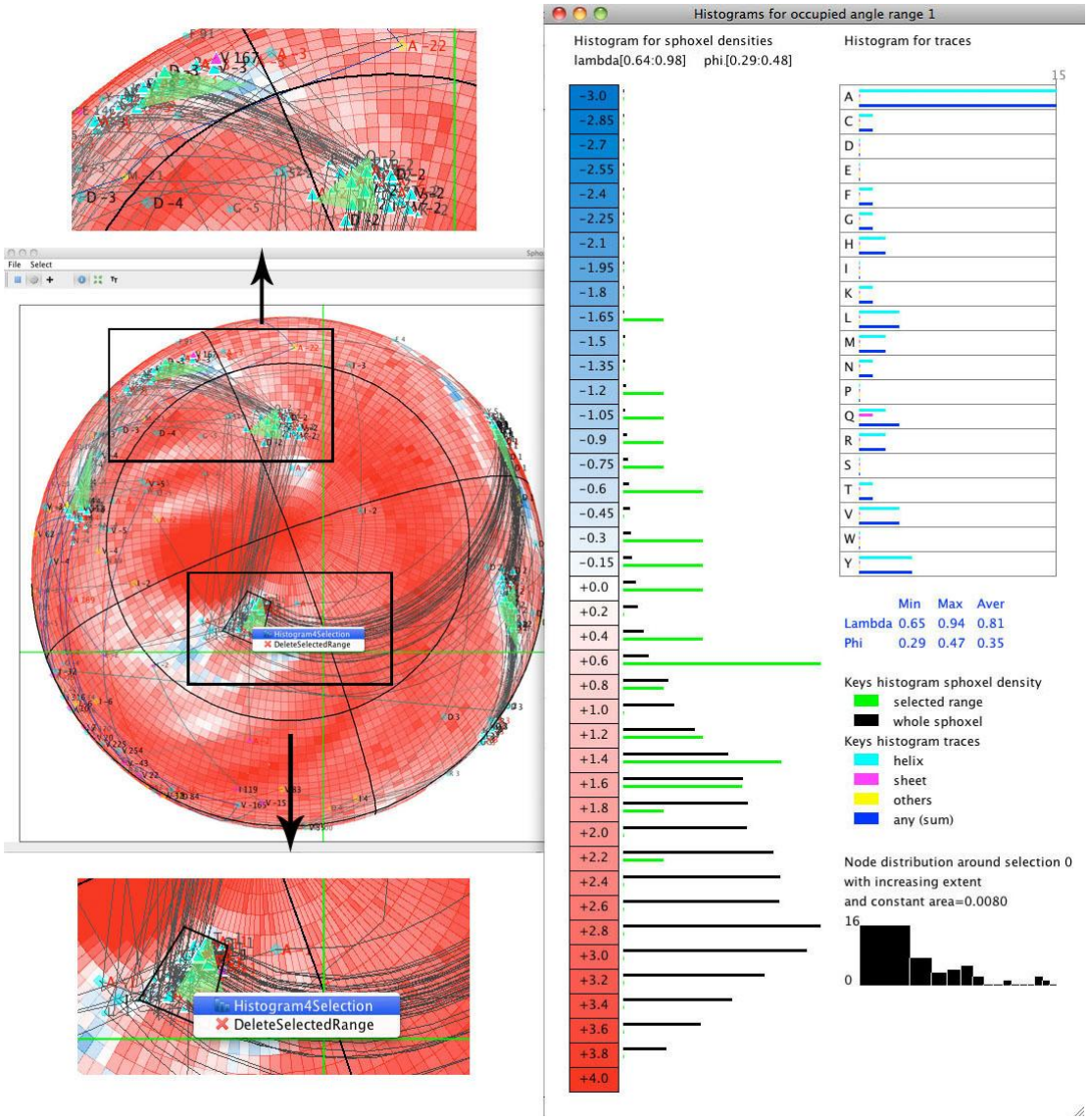


Figure 8: Azimuthal projection of contact potentials between Aspartic Acid and Alanine with a scaled blue-red color-scale. The Neighborhood-Traces are plotted for the template String AxDIAV, excluding the position of the central residue. DBScan was applied, resolved clusters are highlighted and the average direction of outgoing edges from clusters is represented through green arrows. The currently selected dihedral angle ranges is surrounded by a black rectangle. The histogram view for that range is opened via selection within a drop-down menu. Right: Histogram-View for selected angle range. Trace-Histogram shows strong preference towards Alanine contacts with selected range.