Structure Alignment

Homayoun Valafar

Department of Computer Science and Engineering, USC
Status of PDB
Protein Structure Classification Tools

- Many different tools
- Most commonly used three consist of CATH, SCOP, FSSP
- FSSP is based on the DALI algorithm and is fully automated
  - DALI is contact map based method of structure alignment
- SCOP is the most manually curated approach
  - Experts confirm structure classification
- CATH is a semi-automated classification tool
CATH

• **Class**
  - Four classes, alpha, beta, alpha/beta, no structure

• **Architecture**
  - Structural similarity assessed based on orientation of SSE and does not consider connectivity of these SSE
  - This step is conducted manually
  - Assignment is based on previous structural motifs (tim barrel, alpha four helix bundle, beta-propeller)

• **Topology**
  - Grouping is based on the topology of the core domain. Share the same overall shape and connectivity of the SSE

• **Homology**
  - Various levels of sequence identity/similarity
Torsion Angle based structure Alignment (TALI)

• Structure alignment techniques based on contact maps are computationally impractical
• Human intervention methods are limited by human recognition capabilities and are not openly available
• Other hybrid methods such as CATH will fail to identify conserved cores accurately or identify local structural alignment
• There is a need for other structure alignment techniques such as TALI
Structure Alignment Using Torsion Angles

• Two structures can be converted to 2-tuple of angles:
  • Structure 1: (-60, -40) (-60, -50) (-100, 0) …
  • Structure 2: (-60, -40) (-60, -40) (-60, -50) (-100, 10) …

• Structural alignment can be reformulated as sequence alignment.
  • Can use dynamic programming for alignment.

• A meaningful measure of distance is needed.
  • How to measure the distance between two pairs of angles?

• Can use a number of analytical measures of distance.
Protein Structure Alignment: Torsion angle alignment

• Score Function 1:
  • Euclidian distance between \((\phi_i, \psi_i)\) and \((\phi_j, \psi_j)\)

\[
d_{ij}[(\varphi_{a,i}, \psi_{a,i}),(\varphi_{b,j}, \psi_{b,j})] = \sqrt{(\varphi_{a,i} - \varphi_{b,j})^2 + (\psi_{a,i} - \psi_{b,j})^2}
\]
Protein Structure Alignment: Torsion angle alignment (cont.)

• Score Function 2:
  • Convert the density plot of Ramachandran space to $-\log()$.
  • Distance between $(\phi_i, \psi_i)$ and $(\phi_j, \psi_j)$ computed by:

\[
d_{ij}^r = \int_{L} R(l) dl
\]

where $L$ is the path connecting two points in space.
Fundamental Difference Between TALI and Others

• Other programs identify a set of regions in two structures such that:

\[
\left\{ \left( \left\{ R_i^a, R_i^b \right\}, \text{RMSD} \left( \left\{ R_i^a, R_i^b \right\} \right) \right) \right\} = 0
\]

• TALI identifies a set of regions in two structures such that:

\[
\left\{ \left( \left( \sum_{R_i} \text{RMSD} \left( R_i^b \right) \right) \text{is minimized} \right\} R_i^a, R_i^b \right) \right\} + \text{RMSD} \left( \left( R_3^a, R_3^b \right) \right) = 0
\]

• TALI identifies regions of deviation.
Alignment Result 2

2) lcewi-1r4ca

3) lhngb-1a64a
Alignment Result 3: Phylogeny of class II amonoacyl-tRNA synthetase
Why Stop at Torsion Angles

• Residue-to-residue match score:
  • Torsion angles
  • Hydrophobicity
  • Surface accessibility
  • Sequence
  • Distance from protein core
  • Others?
Protein S/R Kinase
Why Not Multiple Structure Alignment?
## Acyl Carrier Proteins

<table>
<thead>
<tr>
<th>Code</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>2FAC</td>
<td>tieervkkiigeqglgv--kqeevtnnasfvedlgadsldtvelvmaleeefdteipdeeae</td>
</tr>
<tr>
<td>1L0I</td>
<td>tieervkkiigeqglgv--kqeevtnnasfvedlgadsldtvelvmaleeefdteipdeeae</td>
</tr>
<tr>
<td>AcpP</td>
<td>diaervkkividhlgv--dadkvvesasfiddlgadsldtvelvmafeeeefgveipddaad</td>
</tr>
<tr>
<td>2JQ4</td>
<td>--natreilakfgqlptpvdttiadeadl-yaaglssfasvqlmlgieeafdiefpdnlln</td>
</tr>
<tr>
<td>1ACP</td>
<td>tieervkkiigeqglgv--kqeevtnnasfvedlgadsldtvelvmaleeefdteipdeeae</td>
</tr>
<tr>
<td>AcpXL</td>
<td>atfdkvadiaetsei--dratitpeshtiddlgidsldfd1divfaidkefgikiplekwt</td>
</tr>
<tr>
<td>2EHS</td>
<td>-leervkeiiaetqglv--ekekitpeakfvedlgadsldvvelimaeeefgieipdedae</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Code</th>
<th>Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>2FAC</td>
<td>458899888898</td>
</tr>
<tr>
<td>1L0I</td>
<td>898008788898</td>
</tr>
<tr>
<td>AcpP</td>
<td>898888889898</td>
</tr>
<tr>
<td>2JQ4</td>
<td>889998898988</td>
</tr>
<tr>
<td>1ACP</td>
<td>898989898989</td>
</tr>
<tr>
<td>AcpXL</td>
<td>898989898989</td>
</tr>
<tr>
<td>2EHS</td>
<td>898889998898</td>
</tr>
</tbody>
</table>

**Computer Science and Engineering**

03/29/10
## Acyl Carrier Proteins

<table>
<thead>
<tr>
<th>PDB Code</th>
<th>Sequence 1</th>
<th>Sequence 2</th>
<th>Score 1</th>
<th>Score 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>2FAC</td>
<td>ek--ittvqaaidyin---g-hq-</td>
<td></td>
<td>880067666667866600060530</td>
<td></td>
</tr>
<tr>
<td>1L0I</td>
<td>ek--mttvqaaidyin---g-hq-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AcpP</td>
<td>ds--iltvgdavkfie---k-aq-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2JQ4</td>
<td>nrksfasikaiedtvklildgkea</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1ACP</td>
<td>ek--ittvqaaidyin---g-hq-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AcpXL</td>
<td>tq---e------vn----------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2EHS</td>
<td>ek--iqtvgdvinylk---e-k--</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2FAC</td>
<td>Hl--11BHbbbbbbbbbb---H-Hl-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1L0I</td>
<td>11--11BHbbbbbbbbbb---H-11-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AcpP</td>
<td>11--11BHbbbbbbbbbb---H-11-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2JQ4</td>
<td>11HHHH1HHHHHHHHHHHHHHHH1HHH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1ACP</td>
<td>11--1111HHHHHHHHHHHHHHHHHHHH1HHH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AcpXL</td>
<td>11---1------1B----------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2EHS</td>
<td>Hl--11BHbbbbbbbbbb---H-H--</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Protein Kinase
Protein kinase

nucleotide binding **********  * ATP

1PME      qvfdvgpynygsnymvcsaydnvkvraikkisp-f-ehqtycqrtrlreikillrf
1O6L      ------dylkllgkgtfgkvlvrekatgryyamkirkeviiaakdevahtvtesrvlqnt
1UNL      ---qkyeklexitegtygtvyfaknrettheivalvkrvrld-morevpsalreicllkel
1GZ8      ---enfekvlikitegtygvykarnktgevvalkkirv-------pstaireisllkel

1PME      11111111BBB1BB11BB11BB1BB1BB111BB1BBBB11-1-11HHHHHHHHHHHHHHHHH1
1O6L      ------BBB1BBB1BB11BB111BB1BBB1BBHHHHHH111HHHHHHHHHHHHHHHHH11
1UNL      ---11BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB111-1111HHHHHHHHHHHHHHHHH111
1GZ8      ---11BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB1BB
# Protein kinase

(ATP) *** * (ATP)

<table>
<thead>
<tr>
<th>Code</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1PME</td>
<td>frheniigindiiraptieqmkdvylvthlmgadlykllktqhlsndhicyflyqilrglk</td>
</tr>
<tr>
<td>1O6L</td>
<td>trhpfiltalkyafqth--drlcfvmeyang-gelffhsrervfteerarfygaeivsale</td>
</tr>
<tr>
<td>1UNL</td>
<td>lkhknivrlhdevlhsd--kkltlvfefdqdlkkyfscngd-ldeivksflfqlkglg</td>
</tr>
<tr>
<td>1GZ8</td>
<td>lnhpnivkldvihte--nklylvfelhqlkkfmdasaltgifplksylfqlqglala</td>
</tr>
</tbody>
</table>

Scores: 8898898788887600788779888788737558888776738878888889898897
Protein kinase

proton acceptor * **** (ATP) * (ATP)

1PME  kyihsanvlhrdlkpnslllntt-dlkicdfglarvadpdhdhtg-fl-teyvatrwy
1O6L  eylhrsrdvvyrdiklenlmdkdghikitdfglcgeisdgatmk-fc-g--tpeylap
1UNL  gfchsrnvlhrdlkpqnllinrnelklanflglarafgipvrcysaevtslwyrppdvl
1GZ8  afchshrvlhrdlkpqnllintegaikladfgralafragvpvrtys-he-v--vtlwyr-

1PME  HHHHH11BB111111HHHBBB111-1BBB111111BB11HHH1111-11-111111HHH1
1O6L  HHHHH1111B111111HHHBB11111BB11111111-11-1--1HHH111
1UNL  HHHHH11BB111111HHHBB11111BB111111111111HHH11HHHH
1GZ8  HHHHH111111111HHHBB11111BB111111111111B--111111-

Scores  788998888898888989887478898889886677677678707707117667773
Mining of Protein Structures

- Structure alignment techniques allow for mining of the PDB
- Answer questions such as:
  - Do all amino acids appear uniformly equally?
  - Do amino acids appear in some structurally relevant regions?
  - Are there structural motifs from which protein structures are made?
Twenty Amino Acids

- Can be categorized based on the following attributes:
  - Charged side chain
  - Polar and uncharged side chain
  - Hydrophobic side chain
  - And others
- Eight amino acids are essential to humans
- Amino acids are precursors to neurotransmitters or nucleotides
- Differ in only the side chain
- Conformation of side chains is determined by $\chi$–angle dihedral
  - Number of dihedral angles depend on the complexity of the side chain
## Basic Information

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>3-Letter</th>
<th>1-Letter</th>
<th>Side chain polarity</th>
<th>Side chain charge (pH 7.4)</th>
<th>Hydropathy Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>Ala</td>
<td>A</td>
<td>nonpolar</td>
<td>neutral</td>
<td>1.8</td>
</tr>
<tr>
<td>Arginine</td>
<td>Arg</td>
<td>R</td>
<td>polar</td>
<td>positive</td>
<td>−4.5</td>
</tr>
<tr>
<td>Asparagine</td>
<td>Asn</td>
<td>N</td>
<td>polar</td>
<td>neutral</td>
<td>−3.5</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>Asp</td>
<td>D</td>
<td>polar</td>
<td>negative</td>
<td>−3.5</td>
</tr>
<tr>
<td>Cysteine</td>
<td>Cys</td>
<td>C</td>
<td>nonpolar</td>
<td>neutral</td>
<td>2.5</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>Glu</td>
<td>E</td>
<td>polar</td>
<td>negative</td>
<td>−3.5</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Gln</td>
<td>Q</td>
<td>polar</td>
<td>neutral</td>
<td>−3.5</td>
</tr>
<tr>
<td>Glycine</td>
<td>Gly</td>
<td>G</td>
<td>nonpolar</td>
<td>neutral</td>
<td>−0.4</td>
</tr>
<tr>
<td>Histidine</td>
<td>His</td>
<td>H</td>
<td>polar</td>
<td>positive (10%), neutral (90%)</td>
<td>−3.2</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>Ile</td>
<td>I</td>
<td>nonpolar</td>
<td>neutral</td>
<td>4.5</td>
</tr>
<tr>
<td>Leucine</td>
<td>Leu</td>
<td>L</td>
<td>nonpolar</td>
<td>neutral</td>
<td>3.8</td>
</tr>
<tr>
<td>Lysine</td>
<td>Lys</td>
<td>K</td>
<td>polar</td>
<td>positive</td>
<td>−3.9</td>
</tr>
<tr>
<td>Methionine</td>
<td>Met</td>
<td>M</td>
<td>nonpolar</td>
<td>neutral</td>
<td>1.9</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>Phe</td>
<td>F</td>
<td>nonpolar</td>
<td>neutral</td>
<td>2.8</td>
</tr>
<tr>
<td>Proline</td>
<td>Pro</td>
<td>P</td>
<td>nonpolar</td>
<td>neutral</td>
<td>−1.6</td>
</tr>
<tr>
<td>Serine</td>
<td>Ser</td>
<td>S</td>
<td>polar</td>
<td>neutral</td>
<td>−0.8</td>
</tr>
<tr>
<td>Threonine</td>
<td>Thr</td>
<td>T</td>
<td>polar</td>
<td>neutral</td>
<td>−0.7</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>Trp</td>
<td>W</td>
<td>nonpolar</td>
<td>neutral</td>
<td>−0.9</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Tyr</td>
<td>Y</td>
<td>polar</td>
<td>neutral</td>
<td>−1.3</td>
</tr>
</tbody>
</table>
Twenty-One Amino Acids

A. Amino Acids with Electrically Charged Side Chains

Positive
- Arginine (Arg)
- Histidine (His)
- Lysine (Lys)
- Aspartic Acid (Asp)
- Glutamic Acid (Glu)

Negative
- Phenylalanine (Phe)
- Phenylalanine (Phe)
- Methionine (Met)
- Tyrosine (Tyr)
- Valine (Val)

B. Amino Acids with Polar Uncharged Side Chains

- Serine (Ser)
- Threonine (Thr)
- Asparagine (Asn)
- Glutamine (Glu)
- Cysteine (Cys)
- Selenocysteine (Sec)
- Glycine (Gly)
- Proline (Pro)

C. Special Cases

- Cysteine (Cys)
- Selenocysteine (Sec)
- Glycine (Gly)
- Proline (Pro)

D. Amino Acids with Hydrophobic Side Chain

- Alanine (Ala)
- Isoleucine (Ile)
- Leucine (Leu)
- Methionine (Met)
- Phenylalanine (Phe)
- Tryptophan (Trp)
- Tyrosine (Tyr)
- Valine (Val)

(Charts and data from NCBI Entrez Gene, 2019)
AA Side Chain

• AA side chains extend from the bb CA atom
• Can be as simple as a H atom
• Side chain atoms are named C-beta, C-gamma, C-delta, C-epsilon, etc.
• Side chain conformation is determined by the $\chi$–angles
<table>
<thead>
<tr>
<th>1</th>
<th>ALA</th>
<th>PSI</th>
<th>165.637</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>LYS+</td>
<td>OMEGA</td>
<td>178.935</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>TRP</td>
<td>OMEGA</td>
<td>-177.200</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>VAL</td>
<td>OMEGA</td>
<td>177.094</td>
</tr>
<tr>
<td>5</td>
<td>CYS</td>
<td>OMEGA</td>
<td>171.732</td>
</tr>
<tr>
<td>6</td>
<td>LYS+</td>
<td>OMEGA</td>
<td>-177.618</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>ILE</td>
<td>OMEGA</td>
<td>176.249</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>CYS</td>
<td>OMEGA</td>
<td>-169.895</td>
</tr>
<tr>
<td>9</td>
<td>GLY</td>
<td>OMEGA</td>
<td>175.765</td>
</tr>
</tbody>
</table>
Occurrence Patterns of AA

- **Glycine**
  - Simplest side chain (a single hydrogen)
  - Less restricted in bb torsion angles
  - Typically is a part of a tight loop
  - Simple SC does not form H-bond and therefore not in SSE

- **Alanine**
  - Next simplest side chain (CH₃)
  - Small hydrophobic
  - Its small size allows presence on the surface of proteins without disrupting its stability
  - Stable exposed hydrophobic patch allows for hydrophobic based interaction
**AA Continued**

- **Cysteine**
  - Small hydrophobic with SC of SH
  - Can form disulfide bridge
  - SH group can coordinate metal binding sites

- **Proline**
  - Its side chain is connected to its bb
  - This restricts its Phi dihedral
  - Does not have amide proton and therefore can not form H-bond
  - Normally in loops and turns
Super Secondary Structural Motifs

- **B-Hairpin**
  - Two residue B-Hairpin
    - Type I' and II'
  - Three or four residue B-Hairpin

The dashed lines indicate main chain hydrogen bonds.

White dots indicate hydrogen bonds.

The main difference between these two turns is the orientation of the peptide group between residues 1 and 2.
Super Secondary Structural Motifs

- Helix-Hairpin
- Helix-turn-Helix
- Beta-Alpha-Beta

A beta-alpha-beta unit
(note the right-handed twist)
Structural Motifs

- Single helix
- Helix-turn-helix
- Four helix bundle
  - Different topologies
- Aligned and Orthogonal B-Sandwich
- B-barrels
- Many more