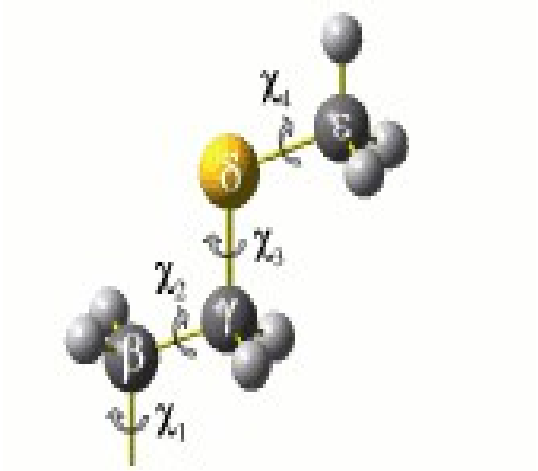
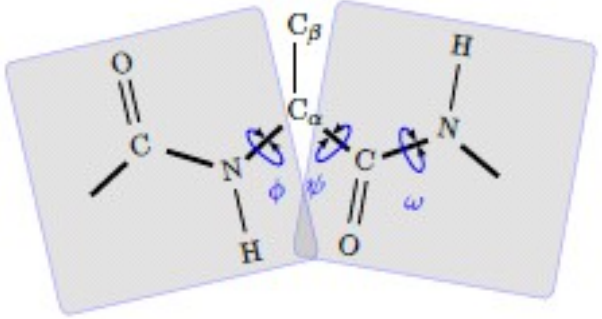
protein.m

Usada para extrair valores de ângulos da conformação de aminoácidos, duma lista de proteína.





Descrição: Decompõe as cadeias de uma proteína em aminoácidos. Registando em ficheiros separados por aminoácido, informação sobre os ângulos dihedrais que definem cada conformação. Para cada aminoácido regista ocorrências de phi, psi, dC_CA, dN_CA, dN_C, dP_plane, ang_N_CA_C, chi1, chi2, chi3, chi4 e TempFactor. A mesma informação é registada para a ocorrência de pares de aminoácidos. Para janelas de três aminoácidos específicos é registado apenas os phi, psi de dada um mais os omega entre eles.

Os ficheiros PDB devem estar na pasta '*proteinData*'

Input: Lista de idêntificações de proteínas do PDB '*pdb1A.tsv*':

Output: Três tipos de ficheiros.

Por amino grava em 'aminoData/<amino>.csv

phi, psi, dC_CA, dN_CA, dN_C, dP_plane, ang_N_CA_C, chi1, chi2, chi3, chi4 e TempFactor.

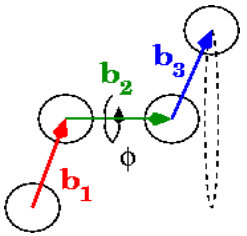
Por par de aminos seguidos grava em 'aminoData/<amino1><amino2>.csv

phi1, psi1, dC_CA1, dN_CA1, dN_C1, dP_plane1, ang_N_CA_C1, chi11, chi12, chi13, chi14, TempFactor1,
phi2, psi2, omega, dC_CA2, dN_CA2, dN_C2, dP_plane2, ang_N_CA_C2, chi21, chi22, chi23, chi24 e TempFactor2

Por cada três aminos seguidos grava em 'rotamer_w3_lib.csv

phi1, psi1, phi2, psi2, omega1, phi3, psi3, omega1

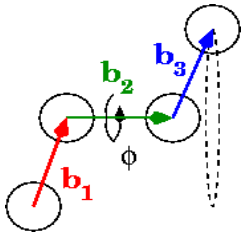
Cada proteína é gravada como uma PDBstructure



[wikipedia]The dihedral angle between two planes relies on being able to efficiently generate a normal vector to each of the planes. One approach is to use the cross product. If A1, A2, and A3 are three non-collinear points on plane A, and B1, B2, and B3 are three non-collinear points on plane B, then UA = (A2−A1) × (A3−A1) is orthogonal to plane A and UB = (B2−B1) × (B3−B1) is orthogonal to plane B. The (unsigned) dihedral angle can therefore be computed with either

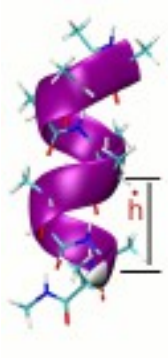
$$\varphi_{AB} = \arccos\left(\frac{U_A \cdot U_B}{|U_A||U_B|}\right) = \arcsin\left(\frac{|U_A \times U_B|}{|U_A||U_B|}\right)$$

protein.m



[wikipedia]The **dihedral angle** between two planes relies on being able to efficiently generate a normal vector to each of the planes. One approach is to use the cross product. If A1, A2, and A3 are three non-collinear points on plane A, and B1, B2, and B3 are three non-collinear points on plane B, then UA = (A2−A1) × (A3−A1) is orthogonal to plane A and UB = (B2−B1) × (B3−B1) is orthogonal to plane B. The (unsigned) dihedral angle can therefore be computed with either

$$\varphi_{AB} = \arccos\left(\frac{U_A \cdot U_B}{|U_A||U_B|}\right) = \arcsin\left(\frac{|U_A \times U_B|}{|U_A||U_B|}\right)$$



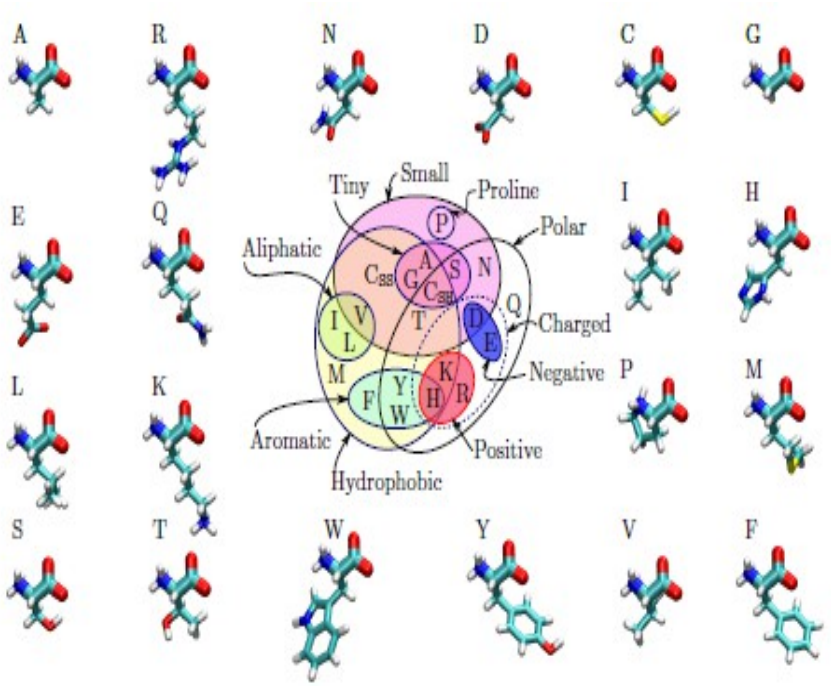

```
HEADER    ALBUMIN-BINDING PROTEIN                 15-JAN-97   1PRB
TITLE     STRUCTURE OF AN ALBUMIN-BINDING DOMAIN, NMR, MINIMIZED
TITLE    2 AVERAGE STRUCTURE
...


REMARK 500 REMARK: NULL
DBREF  1PRB A    1    53  UNP    Q51911   PAB_PEPMA      213    265
SEQRES   1 A   53  THR ILE ASP GLN TRP LEU LEU LYS ASN ALA LYS GLU ASP
SEQRES   2 A   53  ALA ILE ALA GLU LEU LYS LYS ALA GLY ILE THR SER ASP
SEQRES   3 A   53  PHE TYR PHE ASN ALA ILE ASN LYS ALA LYS THR VAL GLU
SEQRES   4 A   53  GLU VAL ASN ALA LEU LYS ASN GLU ILE LEU LYS ALA HIS
SEQRES   5 A   53  ALA
HELIX    1   1 TRP A    5  LEU A    7  5                                   3
HELIX    2   2 ALA A   10  LEU A   18  1                                   9
HELIX    3   3 ASP A   26  LYS A   34  1                                   9
HELIX    4   4 VAL A   38  ILE A   48  1                                  11
CRYST1    1.000    1.000    1.000  90.00  90.00  90.00 P 1           1
ORIGX1      1.000000  0.000000  0.000000        0.00000
ORIGX2      0.000000  1.000000  0.000000        0.00000
ORIGX3      0.000000  0.000000  1.000000        0.00000
SCALE1      1.000000  0.000000  0.000000        0.00000
SCALE2      0.000000  1.000000  0.000000        0.00000
SCALE3      0.000000  0.000000  1.000000        0.00000
ATOM      1  N   THR A   1      23.710  -0.751   1.591  1.00  6.88           N
ATOM      2  CA  THR A   1      22.768  -1.772   1.045  1.00  6.49           C
ATOM      3  C   THR A   1      21.348  -1.202   1.011  1.00  5.61           C
ATOM      4  O   THR A   1      21.084  -0.193   0.387  1.00  5.71           O
ATOM      5  CB  THR A   1      22.777  -3.007   1.944  1.00  6.97           C
ATOM      6  OG1 THR A   1      22.335  -4.133   1.199  1.00  7.22           O
……...
```

protein.m

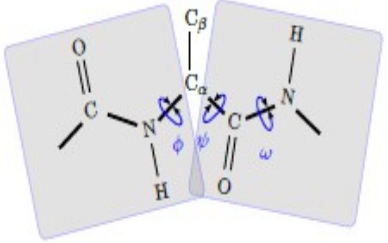
MATLAB structure containing the following fields:

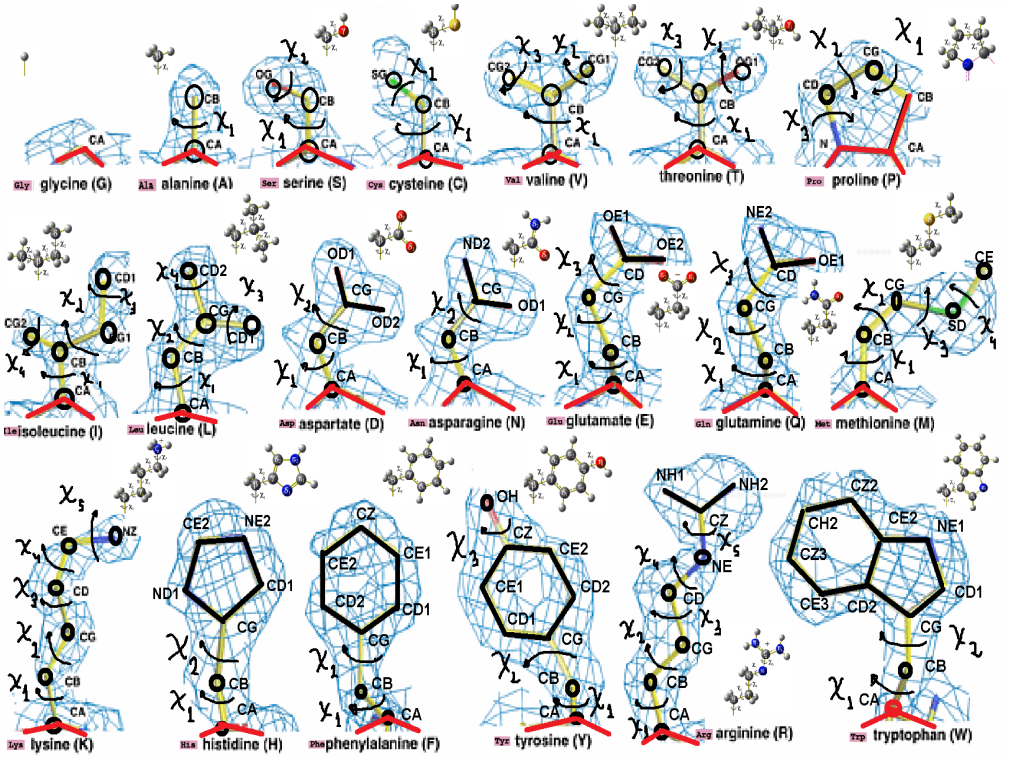
Header, Title, Compound, Source, Keywords, Experiment Data, Authors, Journal, Remark1, Remark2, Remark3, Sequence, HeterogenName, HeterogenSynonym, Formula, Site, Atom, RevisionDate, Superseded, Remark4, Remark5, Heterogen, Helix, Turn, Cryst1, OriginX, Scale, Terminal, HeterogenAtom, Connectivity

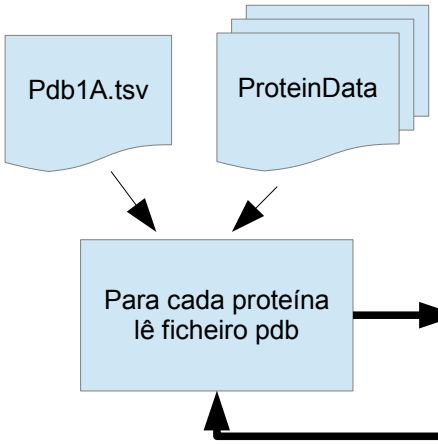


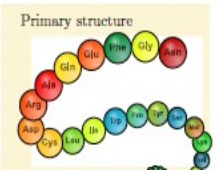
Amino Acid	3-Letter	1-Letter	Mass (Da)	Polarity	Charge (pH 7)	Hydropathy
Alanine	Ala	A	89.98	nonpolar	neutral	1.8
Arginine	Arg	R	174.20	polar	positive	-4.5
Asparagine	Asn	N	132.12	polar	neutral	-3.5
Aspartic acid	Asp	D	133.10	polar	negative	-3.5
Cysteine	Cys	C	121.15	nonpolar	neutral	2.5
Glutamic acid	Glu	E	147.13	polar	negative	-3.5
Glutamine	Gln	Q	146.15	polar	neutral	-3.5
Glycine	Gly	G	75.07	nonpolar	neutral	-0.4
Histidine	His	H	155.16	polar	positive	-3.2
Isoleucine	Ile	I	131.17	nonpolar	neutral	4.5
Leucine	Leu	L	131.17	nonpolar	neutral	3.8
Lysine	Lys	K	146.19	polar	positive	-3.9
Methionine	Met	M	149.21	nonpolar	neutral	1.9
Phenylalanine	Phe	F	165.19	nonpolar	neutral	2.8
Proline	Pro	P	115.13	nonpolar	neutral	-1.6
Serine	Ser	S	105.09	polar	neutral	-0.8
Threonine	Thr	T	119.12	polar	neutral	-0.7
Tryptophan	Trp	W	204.23	nonpolar	neutral	-0.9
Tyrosine	Tyr	Y	181.19	polar	neutral	-1.3
Valine	Val	V	117.15	nonpolar	neutral	4.2

protein.m



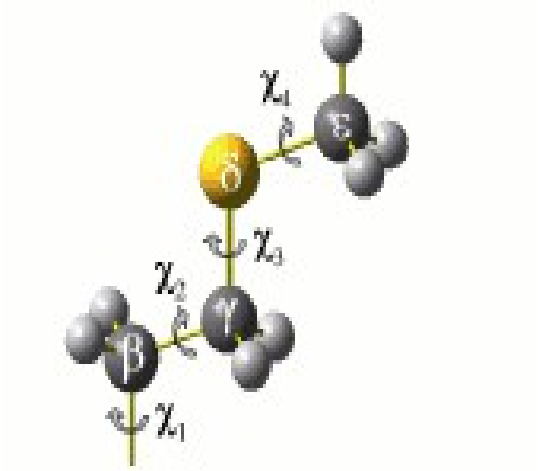

glycine (G)
alanine (A)
serine (S)
cysteine (C)
valine (V)
threonine (T)
proline (P)
isoleucine (I)
leucine (L)
aspartate (D)
asparagine (N)
glutamate (E)
glutamine (Q)
methionine (M)
lysine (K)
histidine (H)
phenylalanine (F)
tyrosine (Y)
arginine (R)
tryptophan (W)

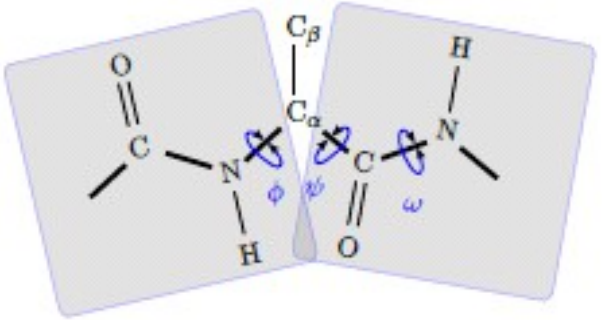

Pdb1A.tsv
ProteinData
Para cada proteína
lê ficheiro pdb
é
RNA ou DNA ou
HYBRID
Não
Para cada cadeia
Calcula na proteína
Ângulos e distâncias
Para cada amino
Append:
aminoData/<amino>.csv
Para cada par
de aminos
consecutivos
Append:
AminoData/
<amino1><amino2>.csv
Para cada três
aminos
consecutivos
Append:
AminoData/
rotamer_w3_lib.csv
Grava PDBstruct
da proteína


Primary structure

Analysis.m

Esta função foi usada para testar estimador de densidade.





Descrição: Analisa informação de cada aminoácido. Calcula matriz com distribuições e gradientes por par de ângulos usando estatística Circular e estimadores de densidade de van Mises. Determina Médias, variâncias, valor máximo e mínimo para a distâncias nos Átomos N CA C na cadeia principal para cada bin. Valor médio Variância, min max para o valor de chi1, chi2, chi3 e chi4 por bin de phi e psi.

Input: Números de bins em que o círculo é decomposto
Ficheiros CVS na pasta 'aminoData/<amini>.csv'

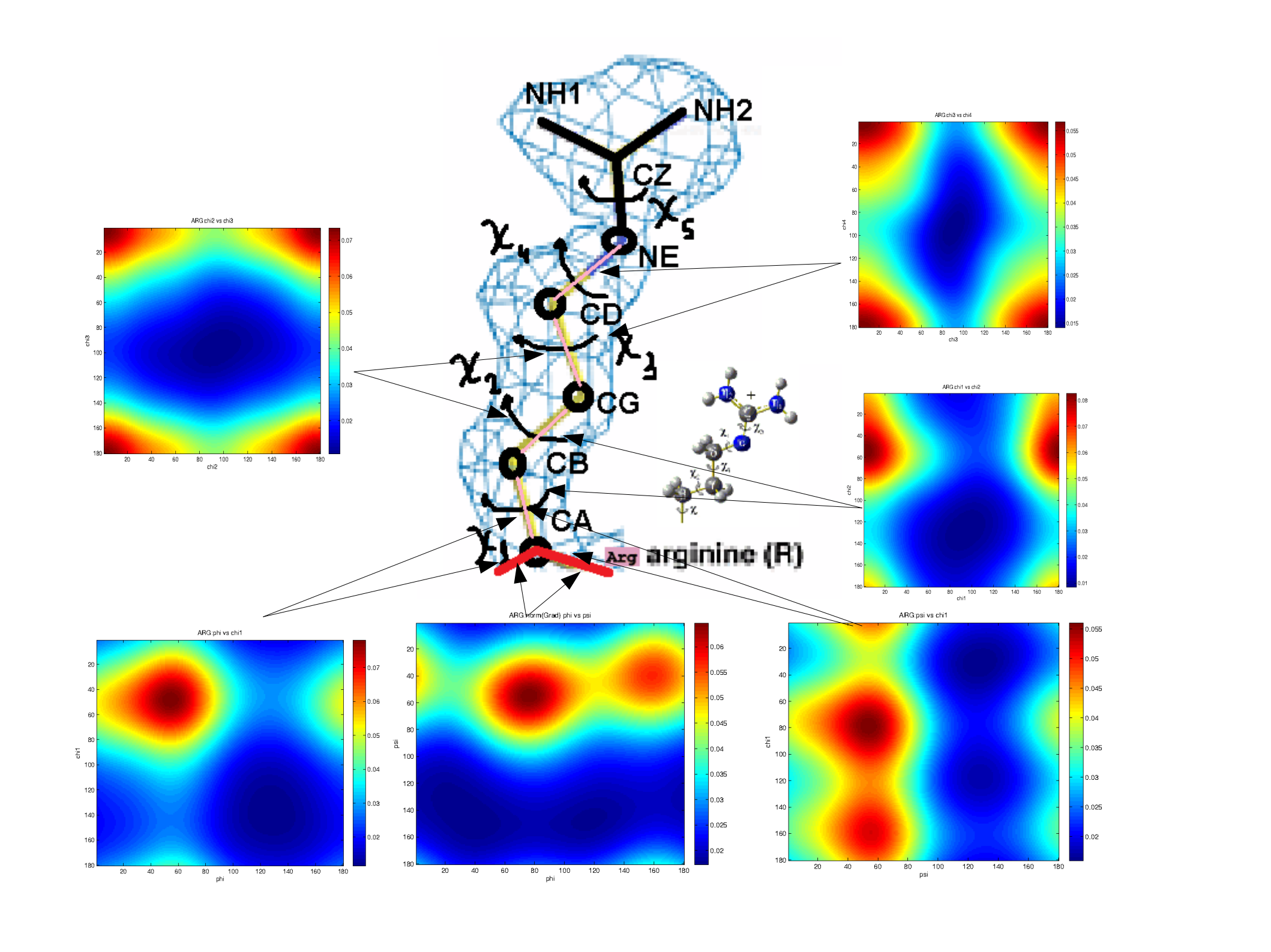
Output: imagens em eps das distribuições entre ângulos por amino
Ramachandran plots:
Phi-psi para 'images/Rama_<amino>.eps'
Phi-chi1 para 'images/Rama_phi_chi1<amino>.eps'
Psi-chi1 para 'images/Rama_psi_chi1<amino>.eps'
Chi1-chi2 para 'images/Rama_chi2<amino>.eps'
Chi2-chi3 para 'images/Rama_chi3<amino>.eps'
Chi3-chs3 para 'images/Rama_chi3<amino>.eps'

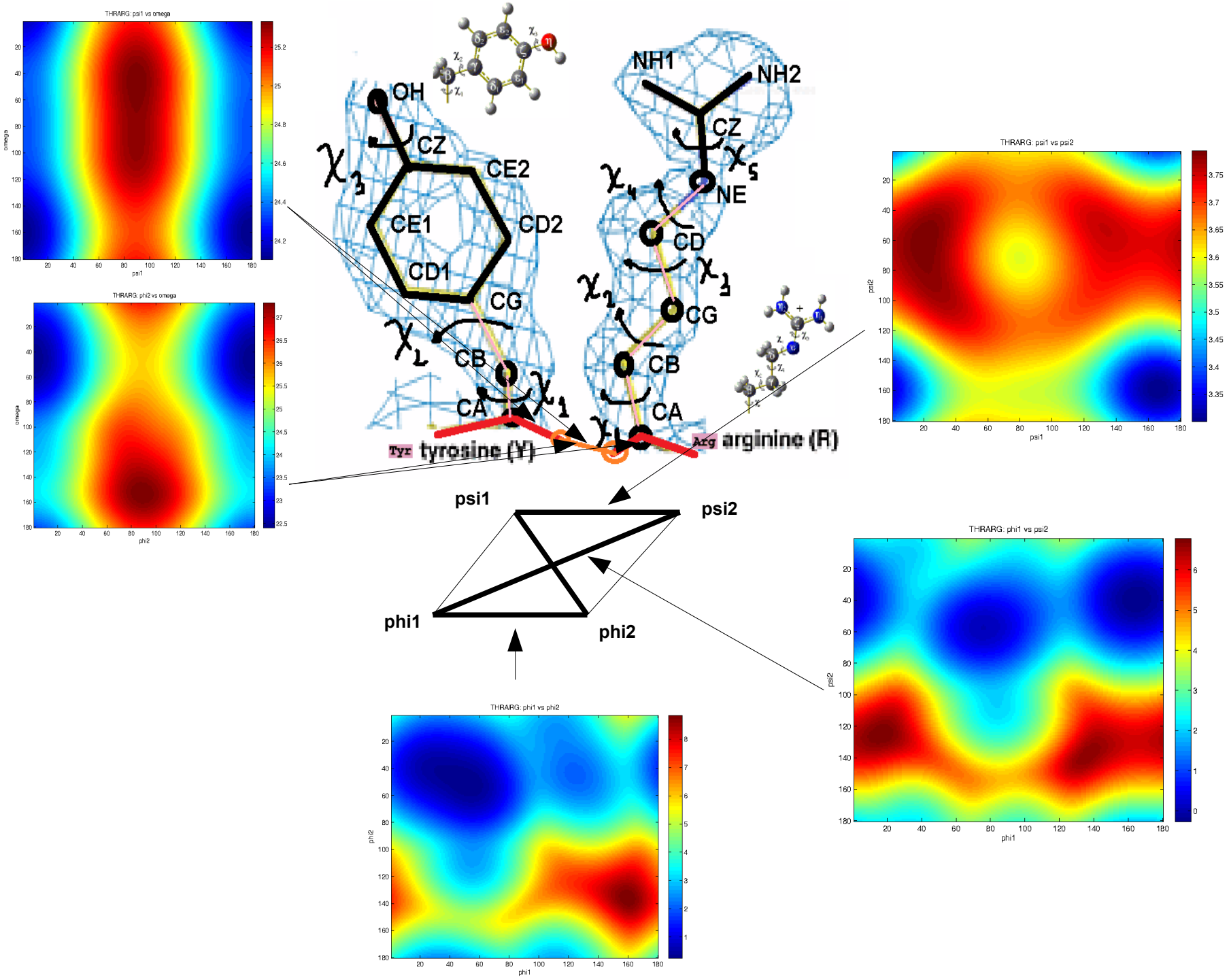
Estrutura com estatísticas por bin (phi,psi):
Distância média, variância, min e max entre:
C e CA, N e CA
Determina ângulos médios para
chi1,chi2,chi3,chi4

Grava estrutura no ficheiro 'binData/<amino>.mat

Amino.dC_CA
Amino.dN_CA
Amino.MeandC_CA
Amino.MeandN_CA
Amino.VardC_CA
Amino.VardN_CA
Amino.mindC_CA
Amino.mindN_CA

Amino.MeanR1 % para chi1
Amino.VarR1
Amino.minR1
Amino.MeanR2 % para chi2
Amino.VarR2
Amino.minR2
Amino.MeanR3 % para chi3
Amino.VarR3
Amino.minR3
Amino.MeanR4 % para chi4
Amino.VarR4
Amino.minR4


NH1
NH2
CZ
NE
CD
CG
CB
CA
Arg
arginine (R)
ARG:chi2 vs chi3
chi3
chi2
ARG:chi3 vs chi4
chi4
chi3
ARG:chi1 vs chi2
chi2
chi1
ARG:phi vs chi1
chi1
phi
ARG:norm(Grad) phi vs psi
psi
phi
ARG:psi vs chi1
chi1
psi


THRARG: psi1 vs omega
omega
psi1
THRARG: phi2 vs omega
omega
phi2
OH
CZ
CE2
CE1
CD2
CD1
CG
CB
CA
NH1
NH2
CZ
NE
CD
CG
CB
CA
Tyr tyrosine (Y)
Arg arginine (R)
THRARG: psi1 vs psi2
psi2
psi1
psi1
psi2
phi1
phi2
THRARG: phi1 vs psi2
psi2
phi1
THRARG: phi1 vs phi2
phi2
phi1