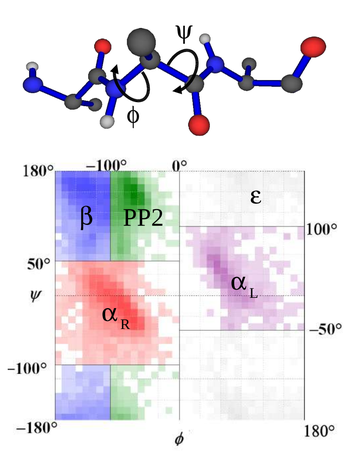
Prediction of Ramachandran basins using fragment libraries

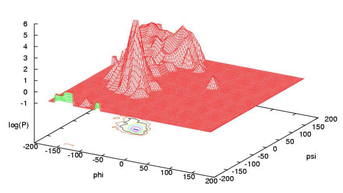
Andrés Colubri[1,2], Abhishek Jha[1,2,3], Tobin R. Sosnick[2,4], Karl F. Freed[1,3], R. Stephen Berry[1,3]
1 Department of Chemistry, 2 Institute of Biophysical Dynamics, 3 The James Franck Institute, 4 Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL 60637
acolubri@uchicago.edu

Overview: We describe a method to construct a library of Ramachandran basin fragments for a given input amino-acid sequence, using BLAST alignments. The input sequence is blasted against a database of known structures, and the aligned matching regions are used to compute Ramachandran basin fragments which are added to the library. This library can be subsequently used to make basin predictions for entire chains, and to guide the conformational search of an ab-initio folding algorithm.

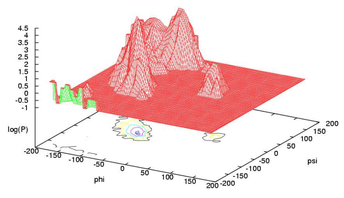
I – Ramachandran basins and local correlations



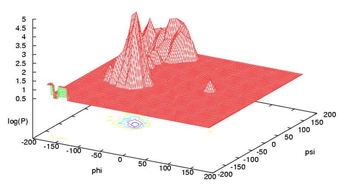
Ramachandran basins used in this study. Number code: 1 = β, 2 = PP2, 4 = α_R, 5 = α_L, 6 = ε (3 is reserved for the upper half of the α_R basin, but this basin has not been introduced yet)



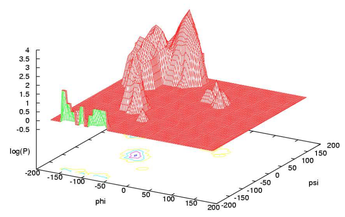
Alanine followed by Alanine



Alanine followed by Glycine



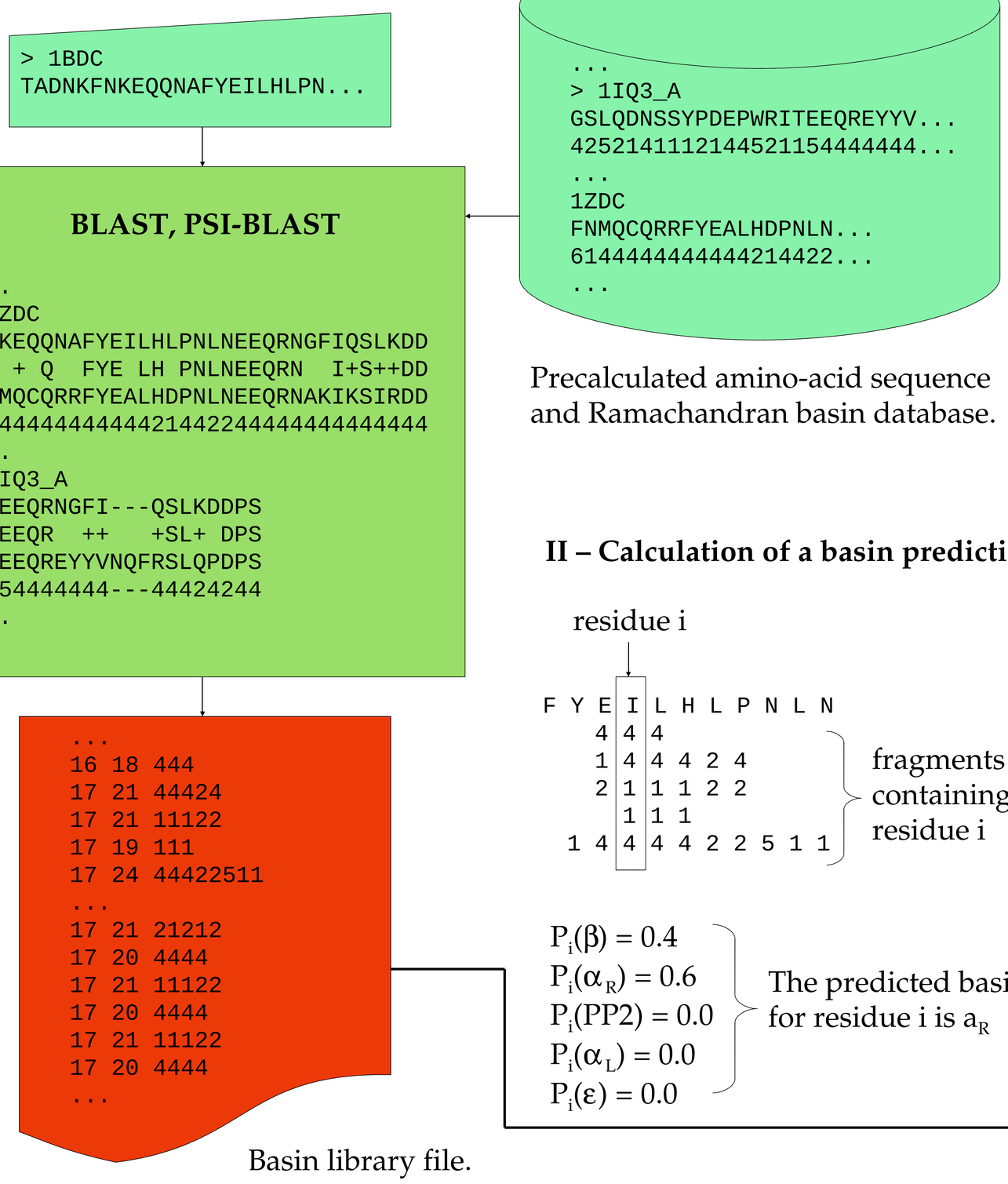
Alanine followed by Histidine



Alanine followed by Proline

The intrinsic occupancies of the basins strongly depend on the first neighbors chemical identity and conformation. We want to capture part of these correlations.

II – Construction of the basin library



Basin library file.

Precalculated amino-acid sequence and Ramachandran basin database.

II – Calculation of a basin prediction

$P_i(\beta) = 0.4$
$P_i(\alpha_R) = 0.6$
$P_i(PP2) = 0.0$
$P_i(\alpha_L) = 0.0$
$P_i(\varepsilon) = 0.0$

The predicted basin for residue i is a_R

III – Results

The percentages of matches between the prediction and the native basins of 50 proteins were computed in three cases: library generated with BLAST, library generated with PSI-BLAST and library generated with the trimer Fragments from the Robetta Fragment server

Code (ID)[2]	BLAST (% hits)	PSI-BLAST (% hits)	Robetta (% hits)	Chain length	Structure topology
1L2Y(A)	44.44	-	50.00	20	α/β
1VII	88.24	82.35	85.29	36	α
1E0L(A)	34.17	40.00	48.57	37	β
1UFI(A)	84.78	89.13	80.43	48	α
1PRB	88.24	88.24	88.24	53	α
1ENH	78.85	80.77	75.00	54	α
3GB1(A)	44.44	66.67	68.52	56	α/β
1JGG(A)	72.73	76.36	76.36	57	α
1UTI(A)	41.82	49.09	67.27	57	β
1SHF(A)	59.65	63.16	63.16	59	β
1OAI(A)	87.72	91.23	89.47	59	α
1BDC	82.76	84.48	84.48	60	α
1R69	78.69	80.33	90.16	63	α
1UCS(A)	38.71	46.77	48.39	64	α/β
1EAY(C)	35.38	41.54	53.85	67	α/β
1DI2(A)	67.16	62.69	80.60	69	α/β
1H64(1)	37.68	50.72	69.57	71	α/β
1PUF(B)	73.24	70.42	84.51	73	α
1DTJ(A)	58.33	68.02	72.22	74	α/β
1UBQ	60.81	90.54	77.03	76	α/β
1VCC	53.33	48.00	65.33	77	α/β
Mean	57.53	65.45	70.83	-	-

IV - Applications

→ Take a look at poster P-C1!

The Ramachandran basins could be used as constrains to reduce the conformational search of a folding algorithm. Our approach is the following:

1.An energy profile is calculated for the current structure.

2. The "hot" residues are selected to change their Ramachandran basins.

3. New basins are chosen from the library, and a 3D structure compatible with those basins is constructed.