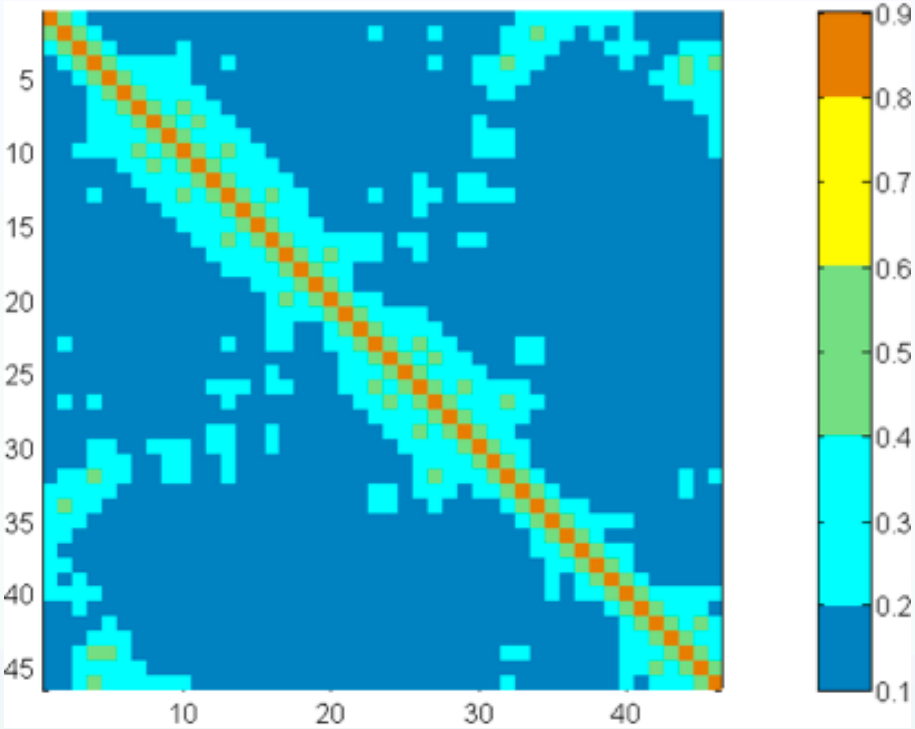
A Spectral Approach to Protein Structure

Aaron Zampaglione
2011 Fall

EIGAs Objective

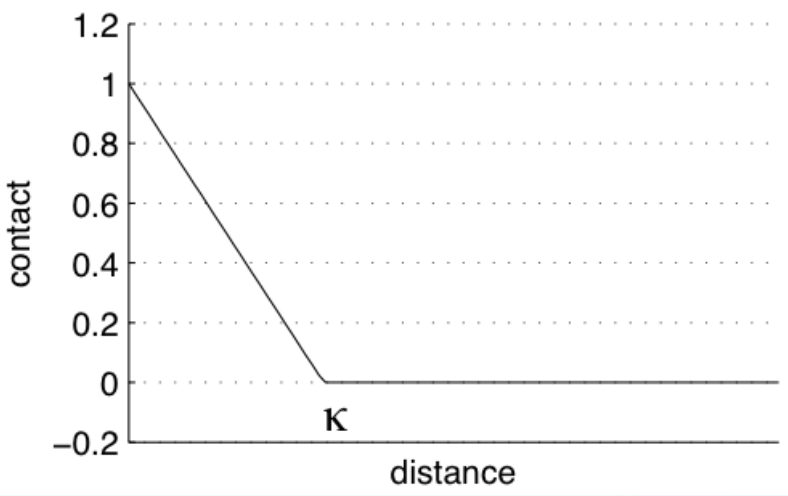
- Structurally compare protein folds.
- Fast.

Contact Matrix



Contact Matrix

$$\sigma(d_{ij}) = \left\{ \begin{array}{ll} 1 - \frac{1}{\kappa} d_{ij} & 0 \le d_{ij} \le \kappa \\ 0 & otherwise \end{array} \right\}$$



Intrinsic Contact Coordinates

$$SVD(CONTACT\_MATRIX) = VDV^T$$

$$R = \sqrt{D}V^T$$

Intrinsic coordinates of residue 4.

R_0	R_1	R_2	R_3	R_4
$\cos(\theta_{0,0})$	…	…		
…	$\cos(\theta_{1,1})$			
…		$\cos(\theta_{2,2})$		
			$\cos(\theta_{3,3})$	
				$\cos(\theta_{4,4})$

Example Intrinsic Matrix for 1FXIa

D_0	D_1	D_2	D_3	D_4	D_5
12.1981	0	0	0	0	…
0	10.4843	0	0	0	…
0	0	10.2397	0	0	…
0	0	0	8.76604	0	…
0	0	0	0	8.27862	…
…	…	…	…	…	…

V_0	V_1	V_2	V_3	V_4	V_5
-0.0610	-0.0848	-0.1495	-0.12	-0.1606	…
0.06328	0.10578	0.18317	0.17656	0.21746	…
-0.0726	-0.078	-0.0983	-0.0303	0.0105	…
0.01813	0.01241	-0.0352	0.02443	0.02211	…
-0.0141	-0.0484	-0.0724	-0.1307	-0.1483	…
…	…	…	…	…	…

Example Intrinsic Matrix for 1FXIa

R_0	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	…
-0.2131	-0.2963	-0.5221	-0.4192	-0.561	-0.358	-0.3596	-0.2868	-0.2078	-0.0855	…
0.2049	0.34252	0.5931	0.57171	0.70412	0.50241	0.40764	0.29229	0.14015	0.04555	…
-0.2325	-0.2496	-0.3146	-0.0971	0.03362	0.3071	0.48553	0.53399	0.45012	0.20938	…
0.05368	0.03676	-0.1043	0.07235	0.06548	0.23138	0.34574	0.3131	0.32754	0.16921	…
-0.0406	-0.1393	-0.2084	-0.3763	-0.4268	-0.4731	-0.4138	-0.337	-0.236	-0.1173	…
0.337	0.40935	0.52426	0.21486	0.07267	-0.3246	-0.4005	-0.4426	-0.275	-0.1214	…
-0.0852	0.02756	0.12779	0.19847	0.16066	-0.1176	-0.35	-0.5326	-0.5777	-0.3437	…
0.42168	0.46185	0.30979	0.28542	0.12126	0.04143	0.02949	-0.0793	-0.0276	-0.0354	…
-0.1046	-0.0322	0.03662	0.01902	-0.0589	-0.011	0.06625	0.15725	0.33221	0.24857	…
0.09955	0.01397	-0.032	-0.1203	-0.0505	-0.1072	-0.0507	-0.0838	-0.1522	-0.13	…
…	…	…	…	…	…	…	…	…	…	…

Protein Fingerprint

$$[\lambda_0, \lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_n]$$

“Standard” DP

$$S_{ij} = \left| \lambda_i - \lambda_j \right|$$

$$M_{i,j} = \min\left(\begin{array}{c} M_{i-1,j} + S_{i-1,j} + 1.0 \\ M_{i-1,j-1} + S_{i-1,j-1} \\ M_{i,j-1} + S_{i,j-1} + 1.0 \end{array} \right)$$

$$init = \begin{array}{c} M_{i,0} = S_{i,0} + i \\ M_{0,j} = S_{0,j} + j \end{array}$$

“Special” DP

$$DIRECTION = \min\begin{pmatrix} M_{i-1,j} + S_{i-1,j} + 1.0 \\ M_{i-1,j-1} + S_{i-1,j-1} \\ M_{i,j-1} + S_{i,j-1} + 1.0 \end{pmatrix}$$

$$IF(DIRECTION == TOP)$$

$$M_{i,j} = M_{i-1,j} + 1.0$$

$$ELSEIF(DIRECTION == LEFT)$$

$$M_{i,j} = M_{i,j-1} + 1.0$$

$$ELSE$$

$$M_{i,j} = M_{i-1,j-1}$$

$$init = \begin{matrix} M_{i,0} = i \\ M_{0,j} = j \end{matrix}$$

“Standard” DP Hard Alignment Results

Protein 1	Protein 2	Our Score	Report’s Score
1FXIa (96)	1UBQ (76)	65	74
1TEN (89)	3HHRb (195)	78	88
3HLAb (99)	2RHE (114)	81	95
2AZAa (129)	1PAZ (120)	93	109
1CEWi (108)	1MOLa (94)	68	88
1CID (177)	2RHE (114)	103	133
1CRL (534)	1EDE (310)	273	304
2SIM (381)	1NSBa (392)	283	339
1BGEb (159)	2GMFa (121)	98	121
1TIE (166)	4FGF (124)	99	120

“Special” DP Hard Alignment Results

Protein 1	Protein 2	Our Score	Report’s Score
1FXIa (96)	1UBQ (76)	74	74
1TEN (89)	3HHRb (195)	89	88
3HLAb (99)	2RHE (114)	95	95
2AZAa (129)	1PAZ (120)	107	109
1CEWi (108)	1MOLa (94)	90	88
1CID (177)	2RHE (114)	112	133
1CRL (534)	1EDE (310)	307	304
2SIM (381)	1NSBa (392)	348	339
1BGEb (159)	2GMFa (121)	119	121
1TIE (166)	4FGF (124)	121	120

Skolnick Results

Fold 1	Fold 2	% Aligned avg.	% Aligned stdv.
TIM B/A Barrel	TIM B/A Barrel	83.23	5.70
TIM B/A Barrel	Ferritin	65.83	4.64
TIM B/A Barrel	Flavodxin	48.40	2.24
TIM B/A Barrel	Cupredoxin	39.60	1.84
TIM B/A Barrel	Mic. Ribo.	40.96	1.67
Ferritin	Ferritin	83.91	7.47
Ferritin	Flavodxin	68.62	3.95
Ferritin	Cupredoxin	57.13	3.45
...	...	...	...

Future

- Implement EIGA.
- Find structural motifs using local (spectral) alignment.
- Compare the motifs using global (spectral) alignment.

References

- Shibberu Y., Holder A. A Spectral Approach to Protein Structure Alignment. Mathematics Department Rose-Hulman Institute of Technology. Terre Haute, IN 47803. 2011. DOI= http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5710873.