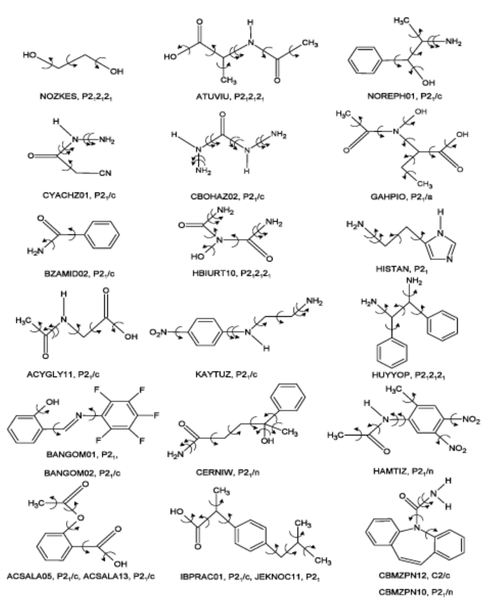
Crystal structure prediction of flexible molecules with genetic algorithms and a standard force field

Julio C. Facelli, [a,b,] Seonah Kim ,[a] Anita M. Orendt,[a] Marta B. Ferraro,[c] Ian Pimienta[a] and Victor Bazterra [a,c]*

[a] Center for High Performance Computing and [b]Department of Biomedical Informatics , University of Utah, 155 South 1452 East Rm 405, Salt Lake City, UT 84112-0190 and [c]Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pab. I (1428), Buenos Aires, Argentina

In this poster we describe our distributed computing framework for crystal structure prediction, MGAC (Modified Genetic Algorithms for Crystal and Cluster Prediction) and its application to predict the structure of flexible molecules using CHARMM with the Generalized Amber Force Field (GAFF). MGAC is capable of performing crystal structure searches for flexible molecules within any space group and with an arbitrary number of molecules in the asymmetric unit. The distributed computing framework includes a series of tightly integrated computer programs for generating the molecule's force field, sampling possible crystal structures using a distributed parallel genetic algorithm, locally minimization of the structures and classifying, sorting and archiving the most relevant ones. Our results indicate the method can consistently find the experimentally known structures of a set of flexible molecules when GAFF reproduces the torsional energetics of the molecule, but unfortunately in some cases GAFF exhibit serious errors in describing this energetics. The following molecules were studied:

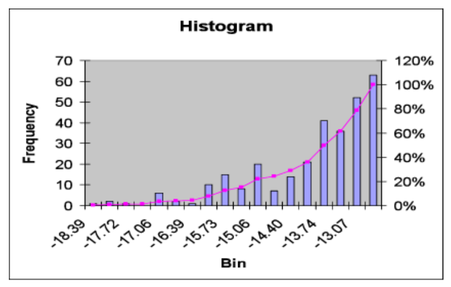


Comparison between predicted and experimental structures

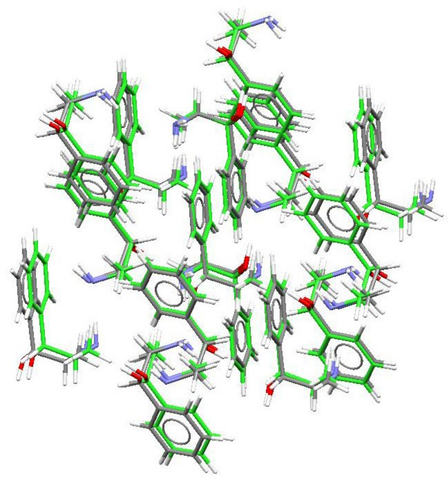
	rank	rms Å	a Å	b Å	c Å	α	β	γ	volume Å^3	space group
NOZKES	78	0.18	4.789	6.763	9.232	90	90	90	299.58	$P2_12_12_1$
(exp.)			5.013	6.915	9.271	90	90	90	321.38	$P2_12_12_1$
NOREPH01	1	0.32	12.447	8.293	7.808	90	104.63	90	779.87	$P2_1/c$
(exp.)			12.507	8.771	8.130	90	106.20	90	856.44	$P2_1/c$
CYACHZ01[a]	196	0.56	7.258	9.309	8.144	90	121.65	90	467.93	$P2_1/c$
(exp.)			7.247	8.678	7.855	90	116.80	90	440.93	$P2_1/c$
CBOHAZ02[b]	110	0.42	3.453	9.196	11.614	90	97.631	90	365.47	$P2_1/c$
(exp)			3.618	8.789	12.487	90	106.43	90	380.85	$P2_1/c$
GAHPIO	1162	1.99	20.150	14.913	5.374	90	90	90	1615.02	$P2_12_12_1$ (Z=8)
(exp.)			14.003	5.425	10.495	90	93.70	90	795.60	$P2_1/a$ (Z=4)
BZAMID02	37	0.64	5.090	4.871	23.31	90	95.58	90	575.15	$P2_1/c$
(exp.)			5.529	5.033	21.343	90	88.73	90	593.77	$P2_1/c$
HBIURT10	106	0.32	11.115	10.386	3.646	90	90	90	420.89	$P2_12_12_1$
(exp.)			10.868	11.698	3.603	90	90	90	458.06	$P2_12_12_1$
HISTAN	2	0.25	7.128	7.253	5.626	90	106.18	90	279.33	$P2_1$
(exp.)			7.249	7.634	5.698	90	104.96	90	304.63	$P2_1$
ACYGLY11[b]	482	0.51	4.895	11.044	10.333	90	101.63	90	547.11	$P2_1/c$
(exp.)			4.859	11.546	14.633	90	138.29	90	546.22	$P2_1/c$
KAYTUZ[b]	22	0.44	10.280	8.807	10.604	90	119.99	90	831.59	$P2_1/c$
(exp.)			10.668	8.958	10.308	90	115.75	90	887.25	$P2_1/c$
HUYYOP	12	0.40	4.738	12.237	18.928	90	90	90	1097.37	$P2_12_12_1$
(exp.)			5.145	12.326	18.536	90	90	90	1175.45	$P2_12_12_1$
BANGOM01	380	0.29	24.563	7.539	5.962	90	90.33	90	1103.33	C2 (Z=4)
(exp.)			12.738	7.263	6.039	90	98.15	90	553.06	$P2_1$ (Z=2)
HAMTIZ	3	0.15	12.521	4.879	17.411	90	100.92	90	1026	$P2_1/c$
(exp)			12.569	4.853	17.266	90	99.16	90	1039.81	$P2_1/n$
ACSALA13[b]	1	0.25	12.371	6.301	11.279	90	112.58	90	811.84	$P2_1/c$
(exp)			12.095	6.491	11.323	90	111.51	90	827.05	$P2_1/c$
CBMZPNXY[b]	11	0.41	27.023	6.478	14.398	90	112.03	90	2336.30	C2/c
(exp) XY=12			26.609	6.926	13.957	90	109.70	90	2421.92	C2/c
	127	0.29	7.490	10.638	29.602	90	90	90	2358.63	P*na*2$_1$ (Z=8)
(exp) XY=10			7.537	11.156	13.912	90	92.86	90	1168.30	$P2_1/n$ (Z=4)

Structures for which the MGAC/GAFF method failed to find the experimental structure

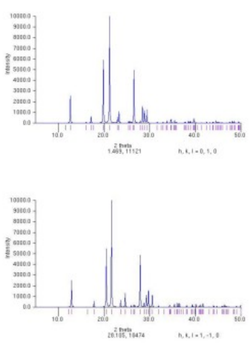
	Reason match not found
ATUVIU	Overestimation of intermolecular HB interaction leading to MGAC structures having smaller unit cell dimensions
BANGOM02	No reason found
CERNIW	Match was found in two dimensions of crystal; possible failure in long range interactions for third crystallographic direction
ASCALA05	Match was found in two dimensions of crystal; possible failure in long range interactions for third crystallographic direction
IBPRAC01	Balance between π···C-H intermolecular and HB intermolecular interactions leading to never seeing herringbone packing found experimentally
JEKNOC11	Balance between π···C-H intermolecular and HB intermolecular interactions leading to never seeing herringbone packing found experimentally



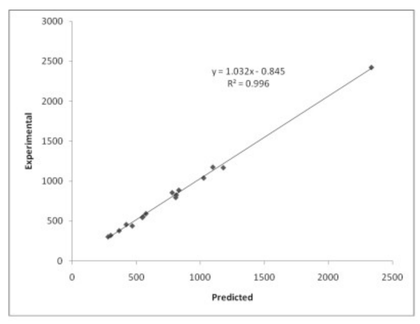
Histogram showing the distribution density of the crystal structures as a function of energy in the MGAC short list for ethylene glycol (NOZKES).



Matching between the predicted (green, rank #1) and experimental structures (gray) of norephedrine (racemic 2-amino-1-phenyl-1-propanol, NOREPH01).



Comparisons between the simulated XRPD patterns of the experimental (top) and predicted (bottom) structures of histamine (HISTAN).



Correlation between the predicted and experimental cell volumes of the compounds studied here. All values in Å^3

Acknowledgements

This work has been partially supported by generous computer time allocations from the NSF TeraGrid award PHY080012N and CHPC allocation on the Arches cluster partially funded by NIH NCRR grant # 1S10RR017214-0. The software for this work used the GAlib genetic algorithm package, written by Matthew Wall at the Massachusetts Institute of Technology. MBF greatly acknowledge financial support from Universidad de Buenos Aires and form the Argentinean CONICET.

* Corresponding author,: Julio.Facelli@utah.edu