

Journal of
Medicinal Chemistry

J. Med. Chem., 1997, 40(8), 1219-1229, DOI:[10.1021/jm960352+](https://doi.org/10.1021/jm960352+)

Terms & Conditions

Electronic Supporting Information files are available without a subscription to ACS Web Editions. The American Chemical Society holds a copyright ownership interest in any copyrightable Supporting Information. Files available from the ACS website may be downloaded for personal use only. Users are not otherwise permitted to reproduce, republish, redistribute, or sell any Supporting Information from the ACS website, either in whole or in part, in either machine-readable form or any other form without permission from the American Chemical Society. For permission to reproduce, republish and redistribute this material, requesters must process their own requests via the RightsLink permission system. Information about how to use the RightsLink permission system can be found at <http://pubs.acs.org/page/copyright/permissions.html>



ACS Publications

MOST TRUSTED. MOST CITED. MOST READ.

Copyright © 1997 American Chemical Society

Supporting Information for:
Selecting Optimally Diverse Compounds from Structure Databases:
A Validation Study of 2D and 3D Molecular Descriptors

Hans Matter

Additional 11 Figures (S1 to S11) with proportions of active structures in active clusters for the IC93 database given for each individual biological class and one table (T1) with an overview over the biological classes from the IC93 database are given.

Supplementary Material Figure Captions:

Figure S1: Proportion of active structures in active clusters for the IC93 database for each individual biological class in comparison to *P0*. The ID of the biological classes is given on the x-axis, while the population is given on the y-axis. 2D fingerprints, atom pair fingerprints and HDisq indices are compared.

Figure S2: Proportion of active structures in active clusters for the IC93 database for each individual biological class in comparison to *P0*. Standard versus modified flexible 3D-fingerprints and 2D-fingerprints, HDisq-indices, modified flexible 3D fingerprints and atom pair fingerprints are compared.

Figure S3: Proportion of active structures in active clusters for the IC93 database for each individual biological class in comparison to *P0*. CoMFA steric fields, Molecular weight and 2D fingerprints, atom-pair distances, modified flexible 3D fingerprints are compared.

Figure S4: Proportion of active structures in active clusters for the IC93 database for each individual biological class in comparison to *P0*. Standard flexible 3D fingerprints, 2D

fingerprints and atom pair fingerprints, 2D fingerprints and modified flexible 3D fingerprints are compared.

Figure S5: Proportion of active structures in active clusters for the IC93 database for each individual biological class in comparison to *P0*. 2D fingerprints, CoMFA steric fields and modified flexible 3D fingerprints are compared.

Figure S6: Proportion of active structures in active clusters for the IC93 database for each individual biological class in comparison to *P0*. CoMFA steric fields and the three different sets of WHIM indices are compared.

Figure S7: Proportion of active structures in active clusters for the IC93 database for each individual biological class in comparison to *P0*. 2D fingerprints, CoMFA steric fields, the second set of WHIM indices and 2D and 3D spatial autocorrelation functions are compared.

Figure S8: Proportion of active structures in active clusters for the IC93 database for each individual biological class in comparison to *P0*. CoMFA steric fields and the three different sets of WHIM indices in combination with 2D fingerprints are compared.

Figure S9: Proportion of active structures in active clusters for the IC93 database for each individual biological class in comparison to *P0*. CoMFA steric fields and 2D fingerprints in combinations with 2D and 3D spatial autocorrelation functions or the second set of WHIM indices together with 3D spatial autocorrelation functions are compared.

Figure S10: Proportion of active structures in active clusters for the IC93 database for each individual biological class in comparison to *P0*. 2D fingerprints alone or in combination with the second set of WHIM indices, 3D spatial autocorrelation functions or modified flexible 3D fingerprints are compared.

Figure S11: Comparison of pairwise biological distances versus various molecular descriptor differences for a dataset consisting of 100 HIV1-protease inhibitors: (a) superposition of molecular structures, (b) molecular steric field, (c) 2D fingerprints, (d) molecular weight.

Table T1: 77 Biological Targets grouped into 55 Classes for 1283 Compounds extracted from the IC93 Database. The 55 classes are denoted by the first part of the class ID running from 100 to 5500. The numbering corresponds to the class IDs used within the text and figures.

100 ANALGESIC_ACTIVITY
200 ANTHELMINTIC_ACTIVITY
300 ANTI-HUMANIMMUNODEFICIENCYVIRUS_ACTIVITY
400 ANTIALLERGIC_ACTIVITY
401 HISTAMINERELEASEINHIBITING_ACTIVITY
402 ANTIANAPHYLACTIC_ACTIVITY
500 ANTIARRHYTHMIC_ACTIVITY
501 CALCIUMANTAGONISTIC_ACTIVITY
600 ANTIASTHMATIC_ACTIVITY
601 BRONCHODILATOR_ACTIVITY
700 ANTIBACTERIAL_ACTIVITY
701 ANTIMICROBIAL_ACTIVITY
702 CELLGROWTHINHIBITING_ACTIVITY
800 ANTICANCER_ACTIVITY
900 ANTICHOLINERGIC_ACTIVITY
1000 ANTICONVULSANT_ACTIVITY
1100 ANTIFEEDANT_ACTIVITY
1200 ANTIFUNGAL_ACTIVITY
1201 FUNGICIDAL_ACTIVITY
1202 HERBICIDAL_ACTIVITY
1203 PHYTOTOXIC_ACTIVITY
1300 ANTIHEPATOTOXIC_ACTIVITY
1400 ANTIHYPERTENSIVE_ACTIVITY
1500 ANTIINFLAMMATORY_ACTIVITY
1600 ANTILEISHMANIAL_ACTIVITY
1700 ANTINEOPLASTIC_ACTIVITY
1800 ANTIOXIDANT_ACTIVITY
1900 ANTISECRETORY_ACTIVITY
2000 ANTITUMOR_ACTIVITY
2100 ANTIULCER_ACTIVITY
2200 ANTIVIRAL_ACTIVITY
2300 BENZODIAZEPINERECEPTORBINDING_ACTIVITY

2400 BETA-ADRENERGIC_ACTIVITY
2500 CHOLECYSTOKININRECEPTORANTAGONISTIC_ACTIVITY
2600 CHOLINERGIC_ACTIVITY
2700 CNS_ACTIVITY
2701 NEUROLEPTIC_ACTIVITY
2702 PSYCHOTROPIC_ACTIVITY
2703 GABARECEPTORAGONISTIC_ACTIVITY
2704 MONOAMINEOXIDASEINHIBITING_ACTIVITY
2705 NEUROMUSCULARBLOCKING_ACTIVITY
2706 OPIATERECEPTORBINDING_ACTIVITY
2707 OPIOIDRECEPTORBINDING_ACTIVITY
2800 CYTOTOXIC_ACTIVITY
2900 DIURETIC_ACTIVITY
3000 DNASYNTHESISINHIBITING_ACTIVITY
3001 PROTEINSYNTHESISINHIBITING_ACTIVITY
3002 RNASYNTHESISINHIBITING_ACTIVITY
3100 DOPAMINED2AGONISTIC_ACTIVITY
3200 ENZYMEINHIBITING_ACTIVITY
3300 ESTROGENRECEPTORBINDING_ACTIVITY
3400 FERTILITYREGULATING_ACTIVITY
3500 GROWTHINHIBITING_ACTIVITY
3600 HEMOLYTIC_ACTIVITY
3700 HISTAMINERELEASING_ACTIVITY
3800 HYPOCHOLESTEROLEMIC_ACTIVITY
3900 IMMUNOSUPPRESSANT_ACTIVITY
4000 INSECTICIDAL_ACTIVITY
4100 INSECTPHEROMONE_ACTIVITY
4200 LEUKOTRIENEANTAGONISTIC_ACTIVITY
4300 LEUKOTRIENERECEPTORANTAGONISTIC_ACTIVITY
4400 NEMATOCIDAL_ACTIVITY
4500 NEUROKININAGONISTIC_ACTIVITY
4600 NEUROKININANTAGONISTIC_ACTIVITY
4700 PLATELETAGGREGATION_ACTIVITY
4701 PLATELETAGGREGATIONINHIBITING_ACTIVITY
4702 THROMBININHIBITING_ACTIVITY
4703 THROMBOXANEAGONISTIC_ACTIVITY

4800 PROTEASEINHIBITING_ACTIVITY

4801 RENININHIBITING_ACTIVITY

4900 RADIOPROTECTIVE_ACTIVITY

5000 SMOOTHMUSCLECONTRACTIONINHIBITING_ACTIVITY

5100 TRYPANOCIDAL_ACTIVITY

5200 TUBULINBINDING_ACTIVITY

5300 ULCEROGENIC_ACTIVITY

5400 VASODILATOR_ACTIVITY

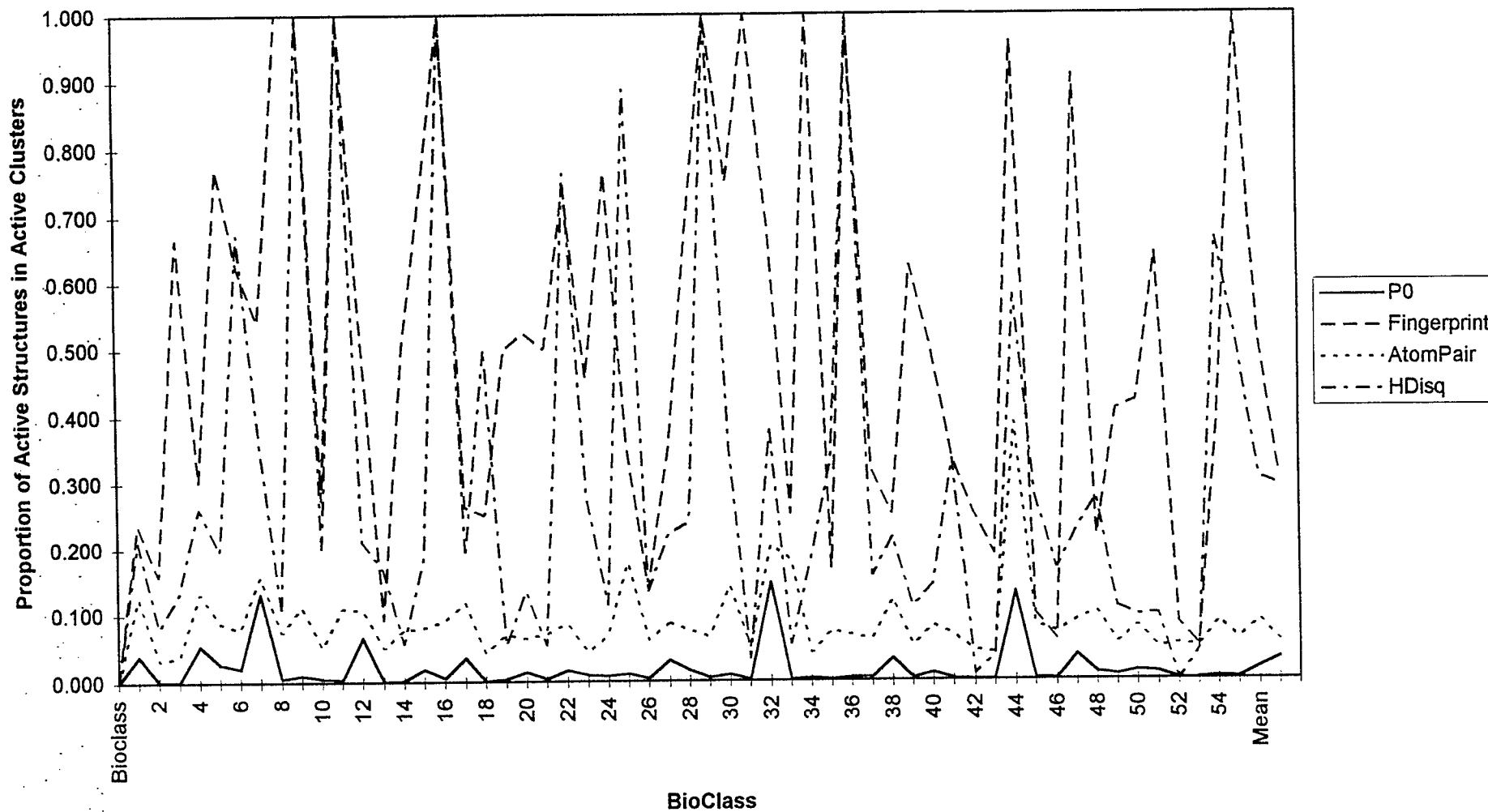
5500 VITAMIND3-LIKE_ACTIVITY

Table T2: Numeric Results from the Comparison of Maximum Dissimilarity Selection versus Random Selection for Compound Subsets from 60 to 500 Structures. ^{a)}

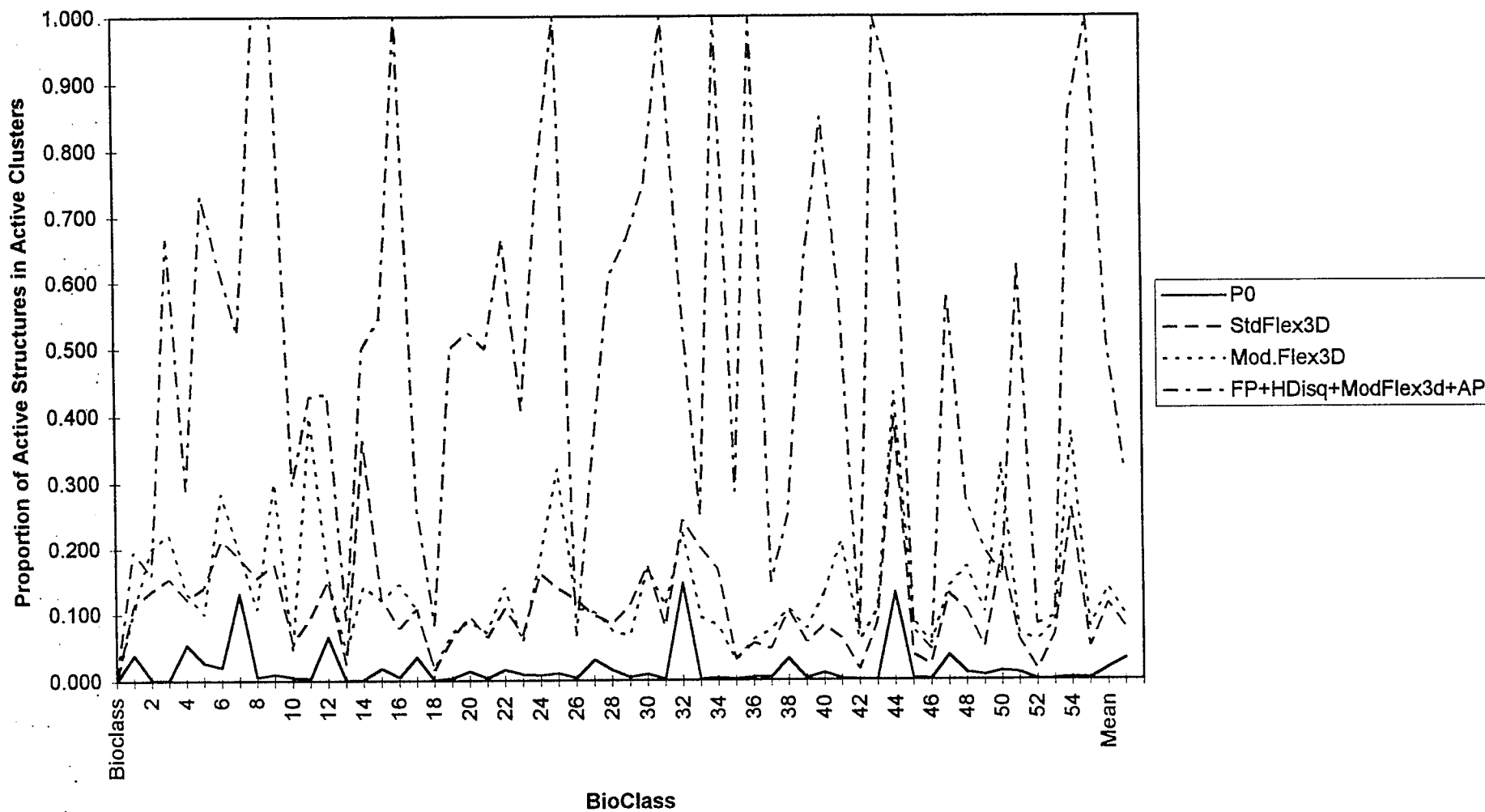
Cpds#	Tan_mean diss	Tan_std diss	Tan_max diss	Class_sel diss	Percent diss	Tan_mean random	Tan_std random	Tan_max random	Class_sel random	Percent random
60	0.31	0.03	0.34	30	54.54	0.65	0.23	1	26	47.27
80	0.34	0.04	0.38	39	70.90	0.67	0.22	1	26	47.27
100	0.37	0.04	0.42	42	76.36	0.71	0.2	1	26	47.27
120	0.4	0.05	0.47	47	85.45	0.71	0.22	1	37	67.27
140	0.43	0.06	0.5	47	85.45	0.76	0.22	1	34	61.81
160	0.46	0.07	0.56	49	89.09	0.77	0.2	1	38	69.09
180	0.49	0.08	0.6	47	85.45	0.79	0.19	1	40	72.72
200	0.52	0.08	0.63	49	89.09	0.79	0.2	1	40	72.72
220	0.55	0.09	0.66	49	89.09	0.81	0.18	1	35	63.63
240	0.57	0.09	0.68	49	89.09	0.81	0.19	1	40	72.72
260	0.6	0.1	0.72	49	89.09	0.82	0.18	1	41	74.54
280	0.62	0.1	0.73	51	92.72	0.84	0.18	1	44	80.00
300	0.64	0.1	0.74	52	94.54	0.84	0.17	1	49	89.09
320	0.66	0.1	0.77	52	94.54	0.85	0.17	1	42	76.36
340	0.67	0.1	0.78	52	94.54	0.85	0.16	1	46	83.63
360	0.69	0.1	0.79	53	96.36	0.85	0.17	1	46	83.63
380	0.71	0.1	0.8	55	100.00	0.86	0.16	1	43	78.18
400	0.72	0.1	0.82	55	100.00	0.86	0.16	1	48	87.27
420	0.73	0.1	0.83	54	98.18	0.87	0.15	1	46	83.63
440	0.74	0.1	0.84	54	98.18	0.86	0.15	1	50	90.90
460	0.75	0.1	0.85	55	100.00	0.87	0.14	1	48	87.27
480	0.76	0.1	0.86	55	100.00	0.88	0.14	1	48	87.27
500	0.77	0.1	0.87	55	100.00	0.88	0.14	1	48	87.27

^{a)} Tan_mean: mean Tanimoto coefficient for maximum dissimilarity (diss) or random selected (random) subset; Tan_max: maximum pairwise Tanimoto coefficient; Class_sel: Number of selected biological targets; Percent: percentage of selected biological classes (from 55 classes).

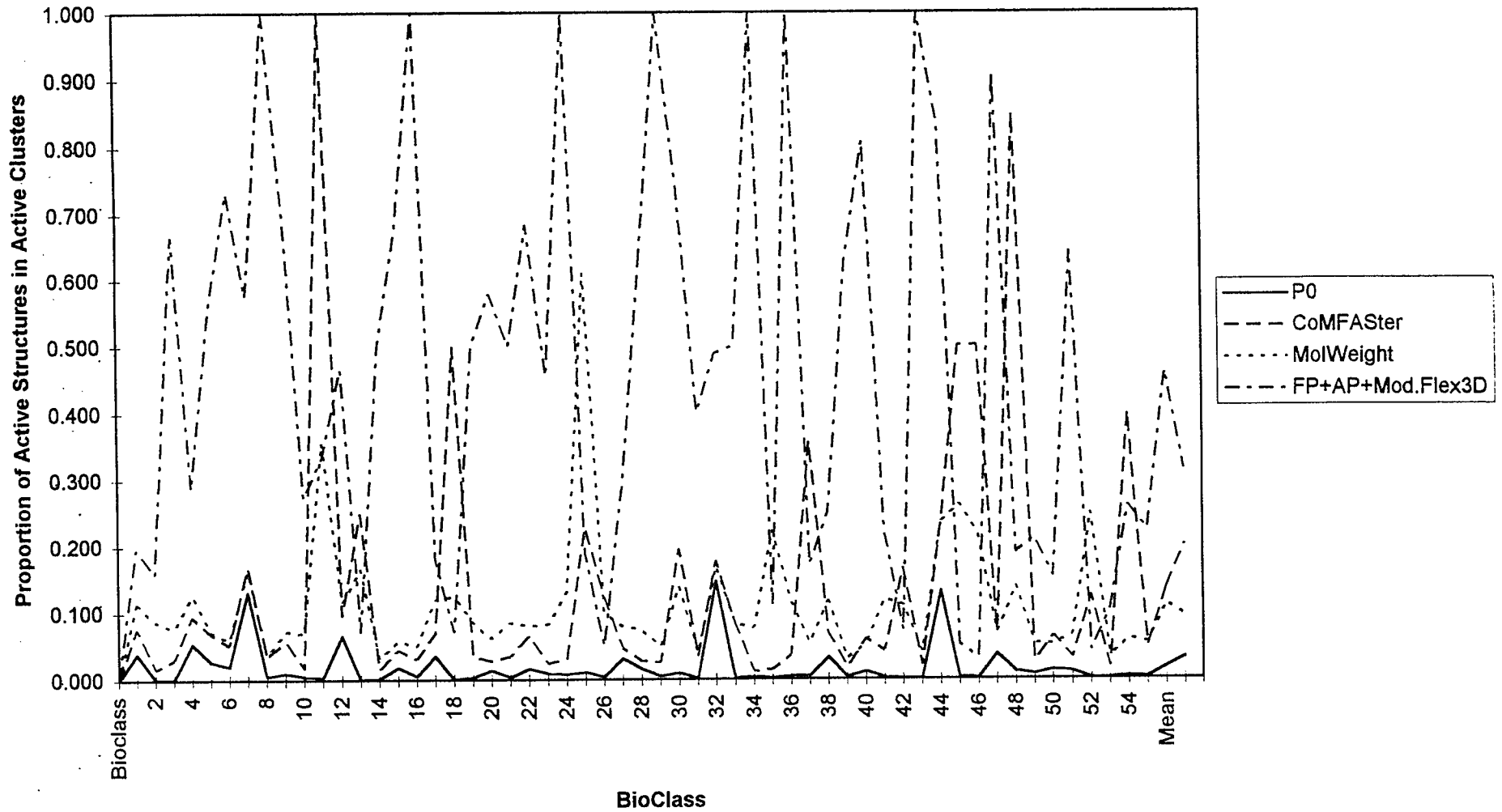
Proportion of Active Structures in Active Clusters for IC93 (I)



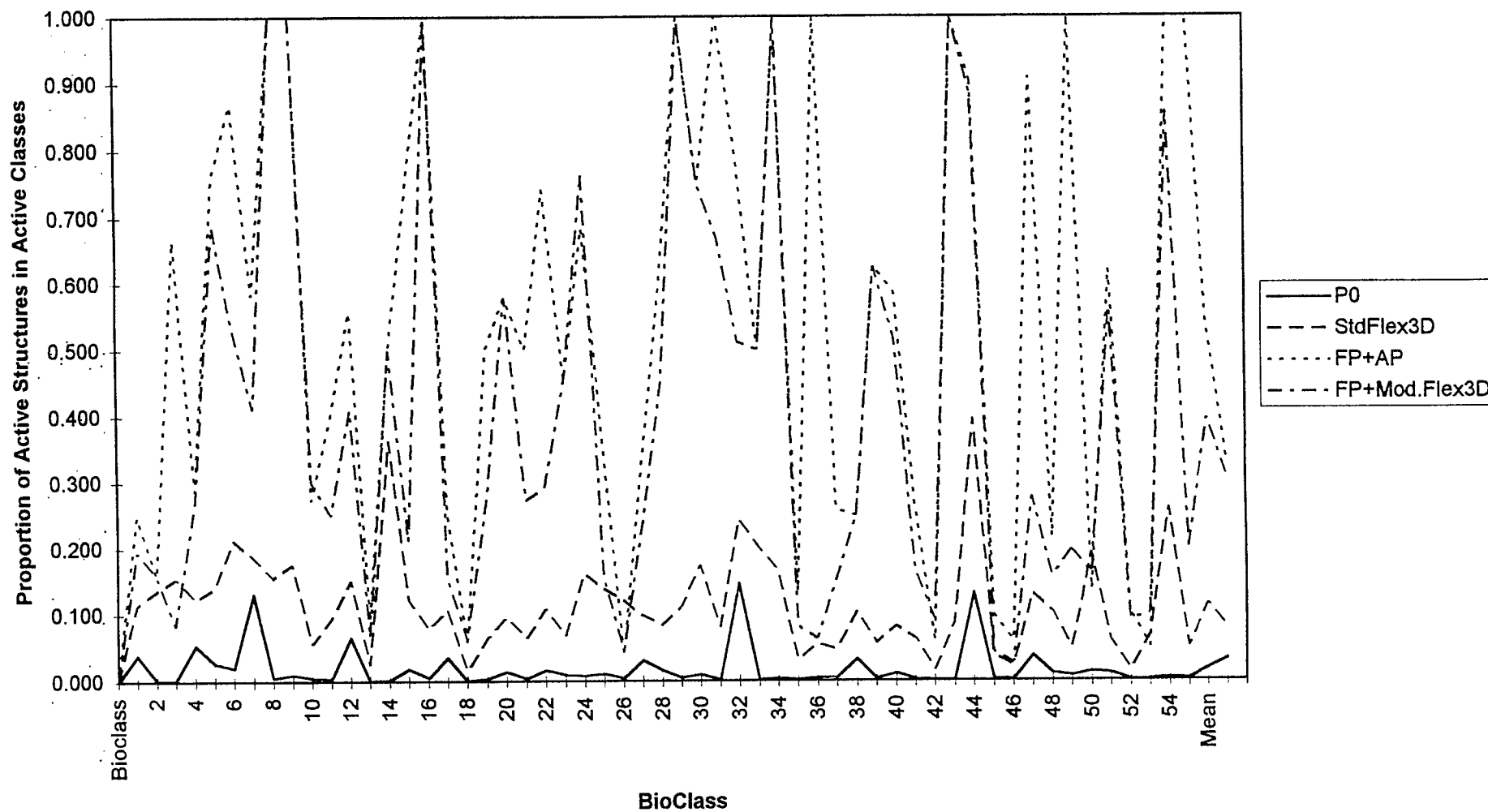
Proportion of Active Structures in Active Clusters for IC93 (II)



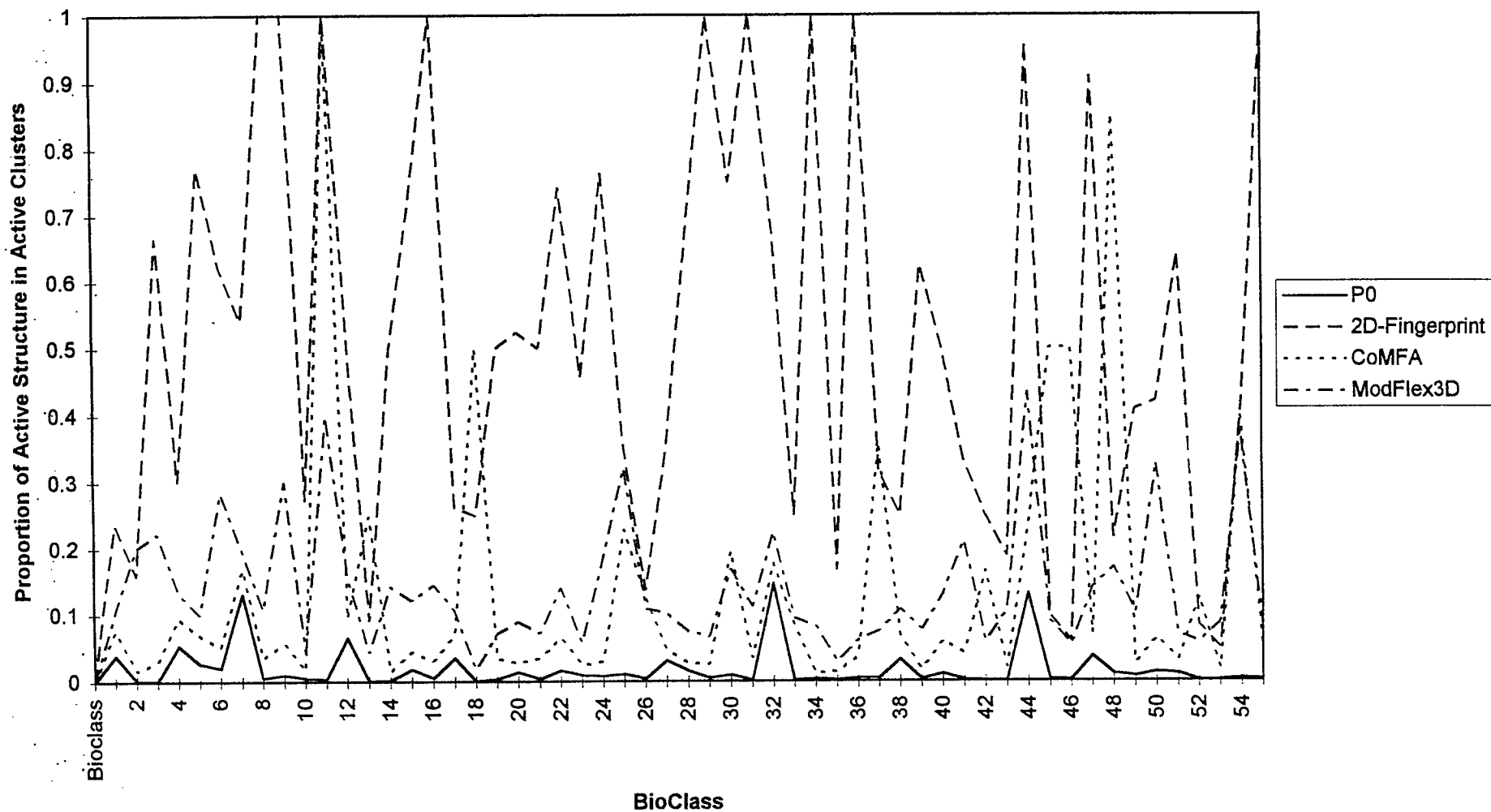
Proportion of Active Structures in Active Clusters for IC93 (III)



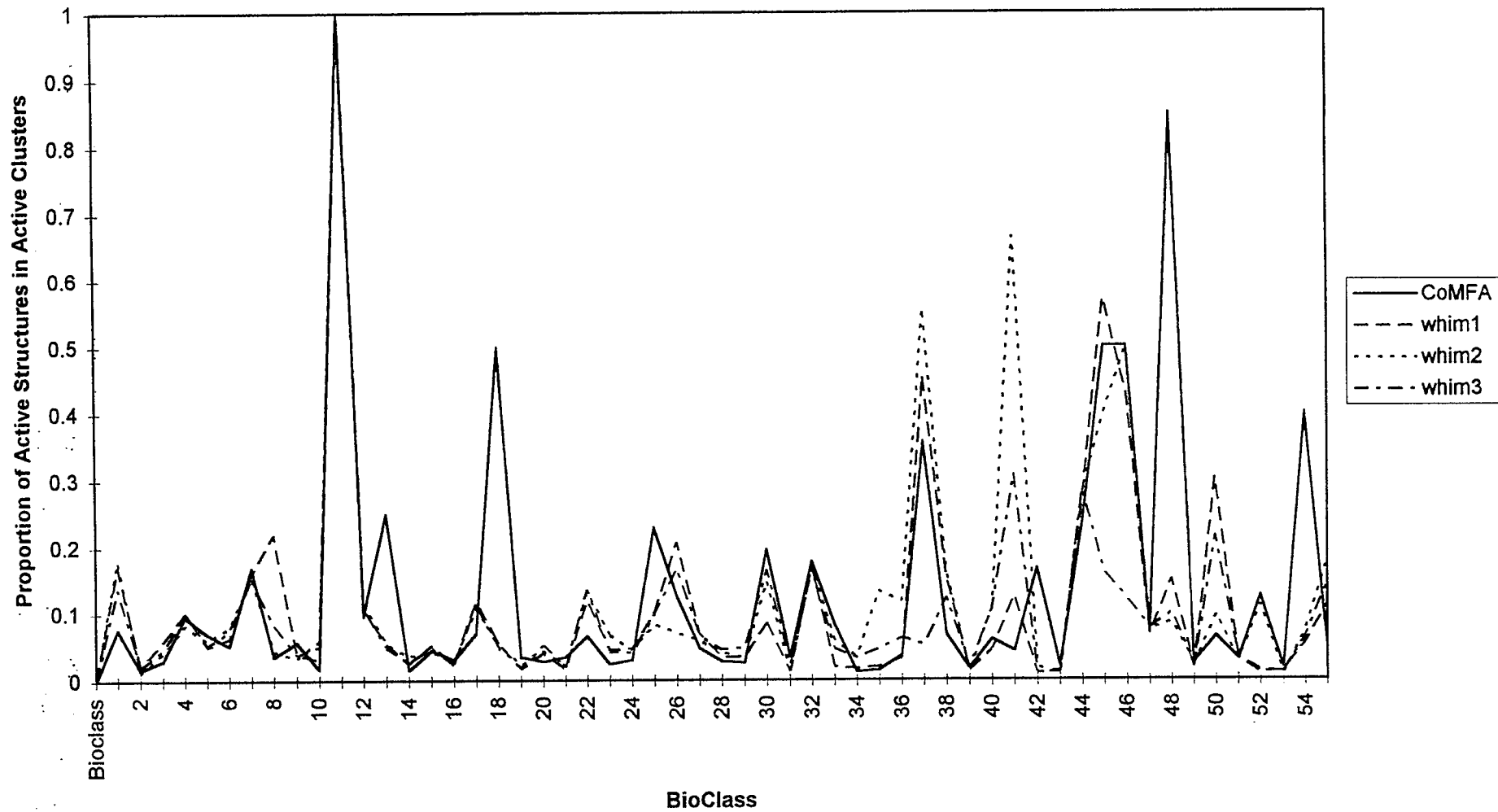
Proportion of Active Structures in Active Clusters for IC94 (IV)



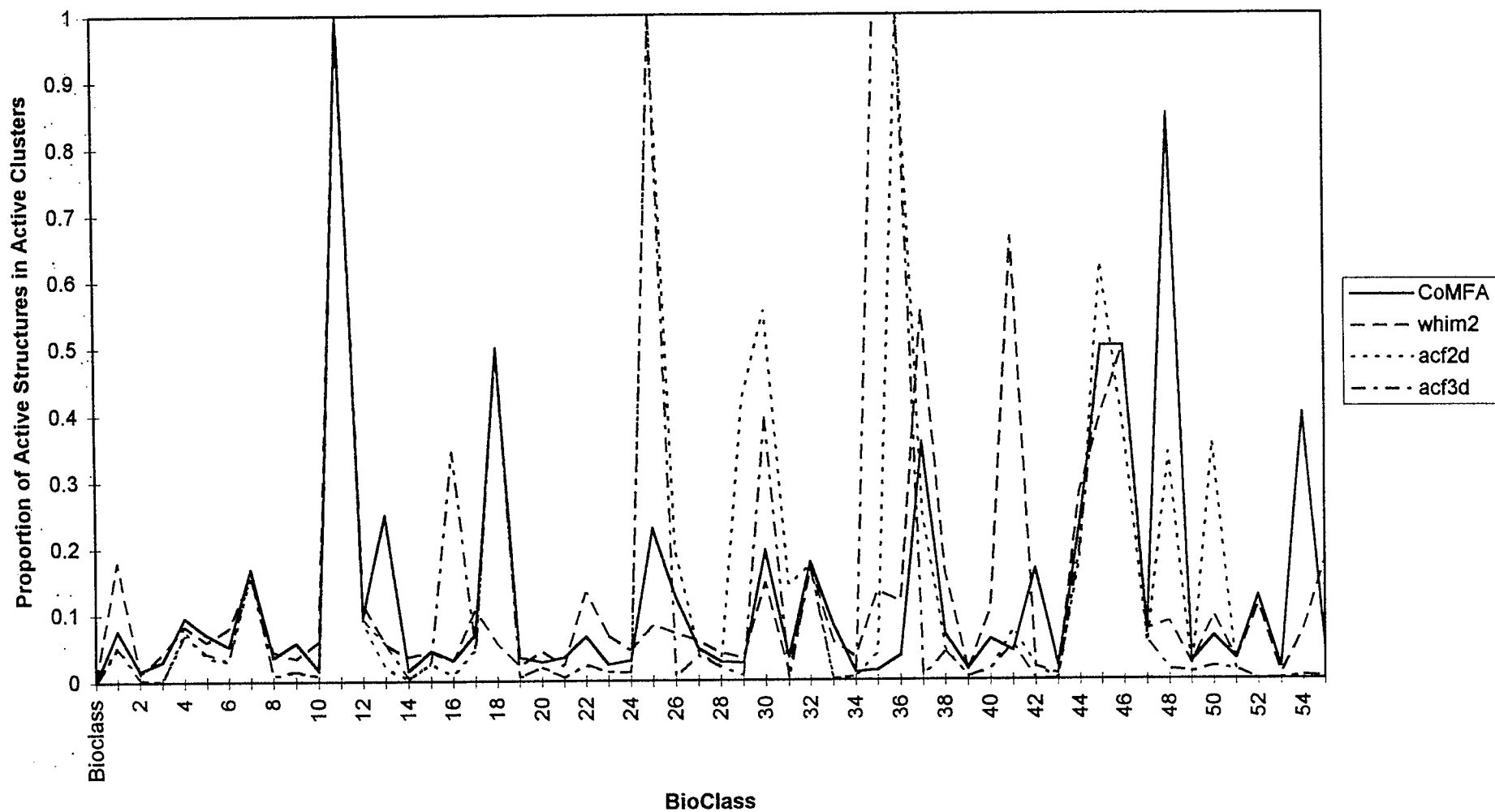
WHIM Indices: Proportion of Active Structures in Active Clusters for IC93 (I)



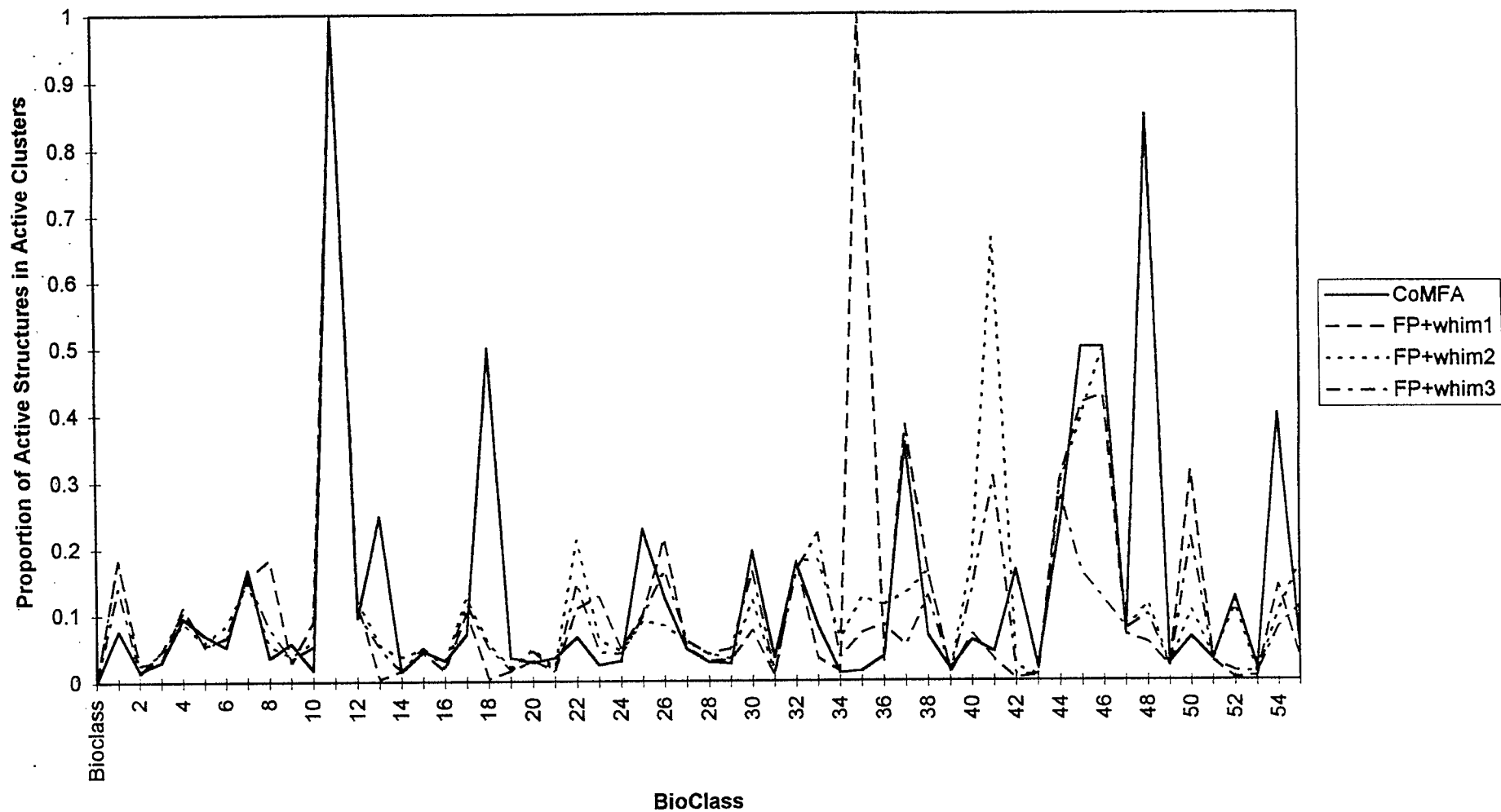
WHIM-Indices: Proportion of Active Structures in Active Clusters for IC93 (II)



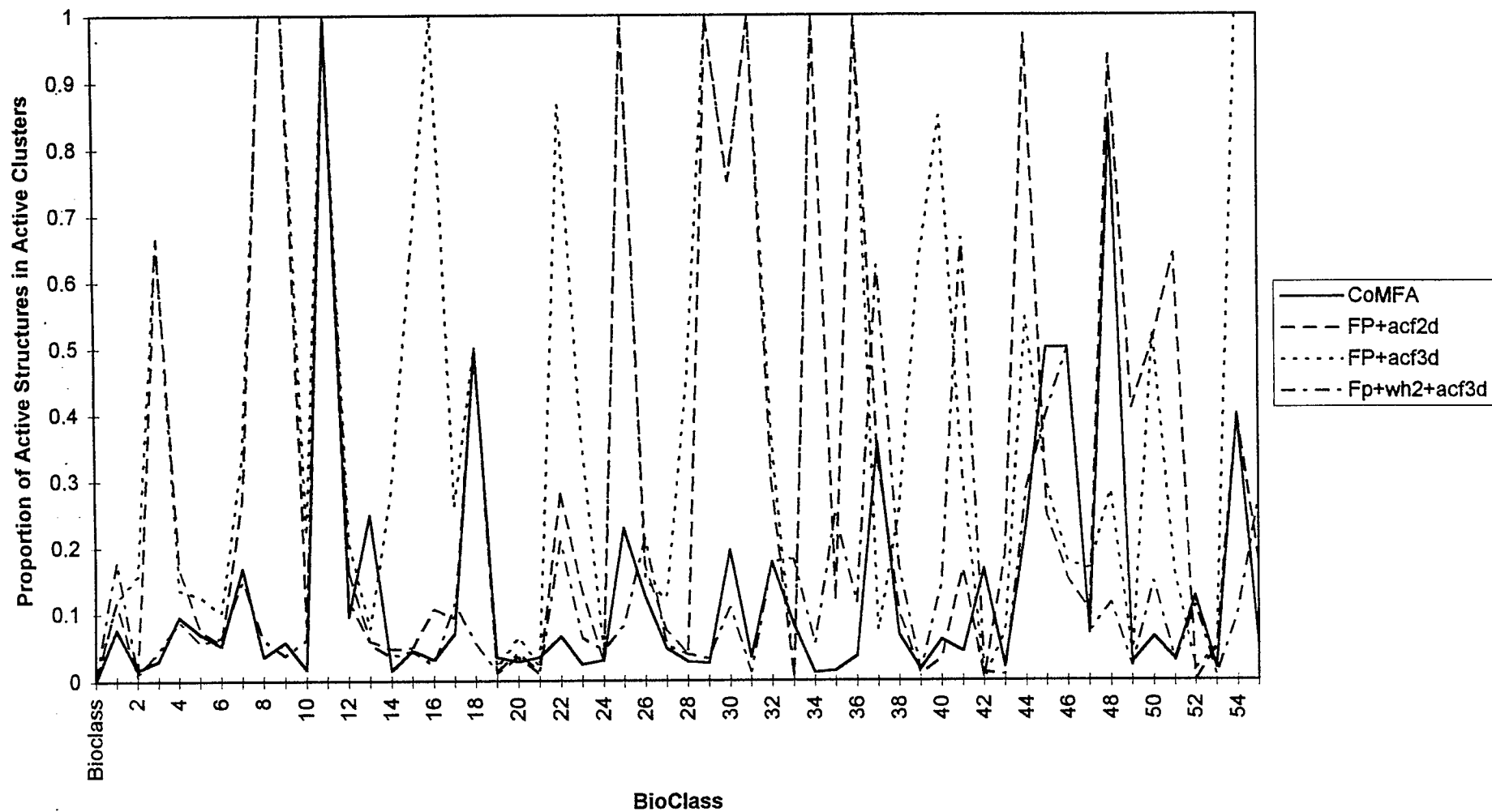
WHIM-Indices: Proportion of Active Structures in Active Clusters for IC93 (III)



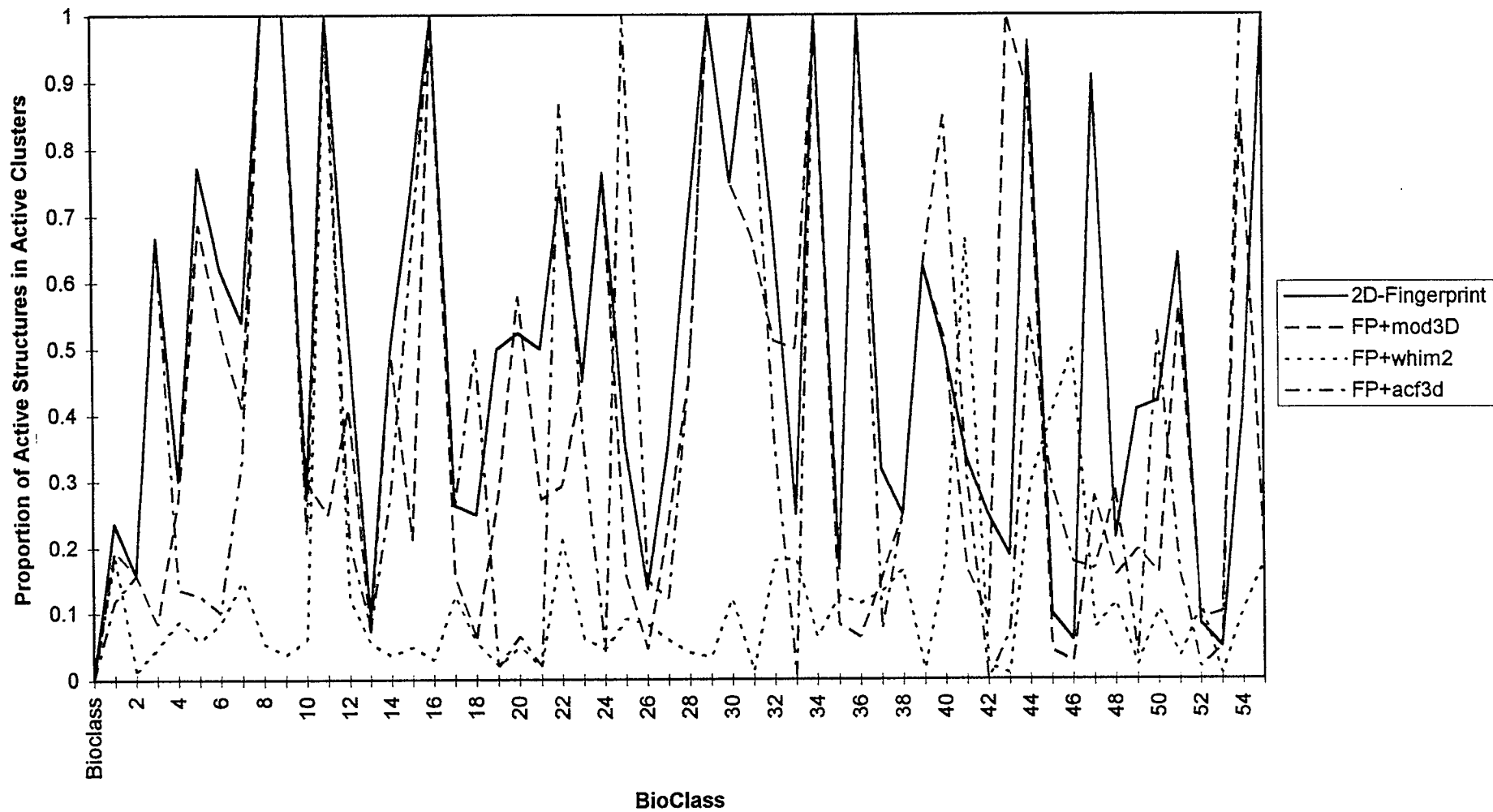
WHIM-indices: Proportion of Active Structures n Active Clusters for IC93 (IV)



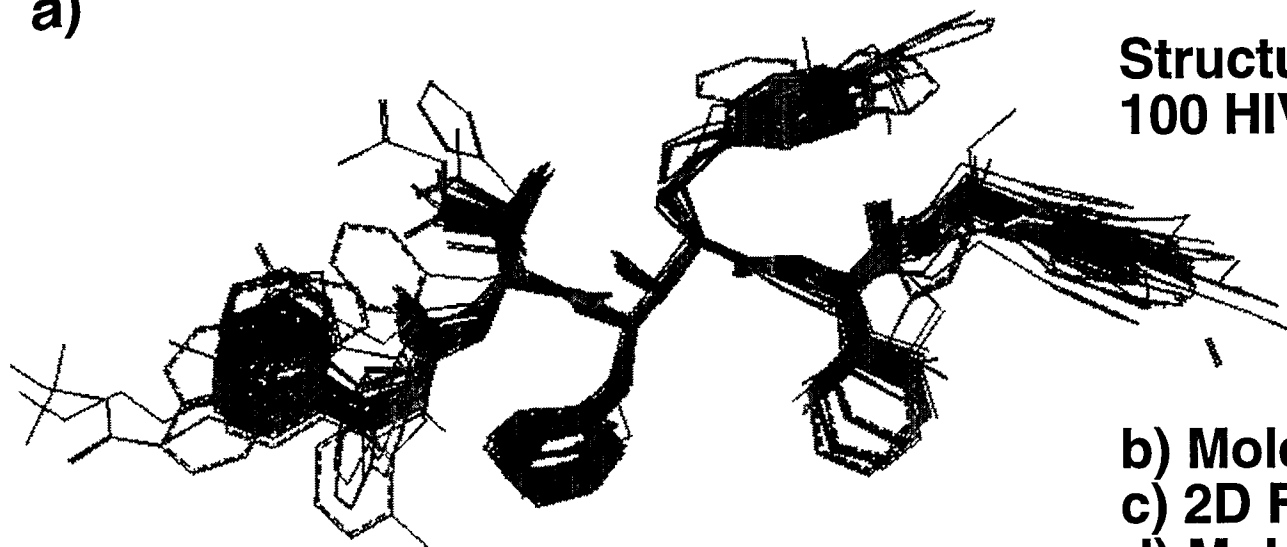
WHIM-Indices: Proportion of Active Structures in Active Clusters for IC93 (V)



Proportion of Active Structures in Active Clusters (Best Descriptors) for IC93 (VI)



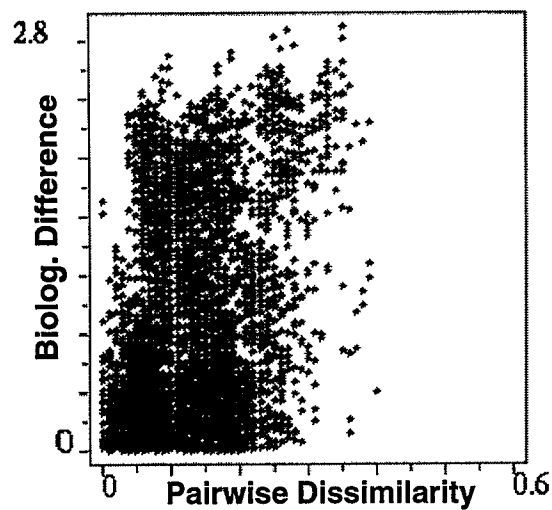
a)



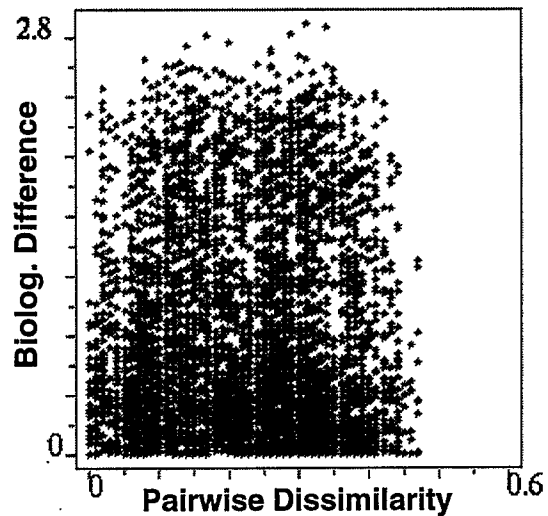
**Structural Diversity of
100 HIV1-Protease Inhibitors:**

- b) Molecular steric fields (CoMFA)**
- c) 2D Fingerprints**
- d) Molecular Weight**

b)



c)



d)

