

Characterizing Activity Landscapes Using an Information-Theoretic Approach

Veer Shanmugasundaram & Gerry Maggiora

Computer-Aided Drug Discovery

Pharmacia Corporation, Kalamazoo, MI

What are Activity Landscapes?

- Activity landscapes are abstract surfaces drawn on chemistry space containing compounds where the height represents biological activity.
- Gentle rising hills of activity represent smooth landscape where small structural changes produce gradual changes in activity.
- Rough activity landscapes are characterized by cliffs where small changes in structure lead to large changes in activity.

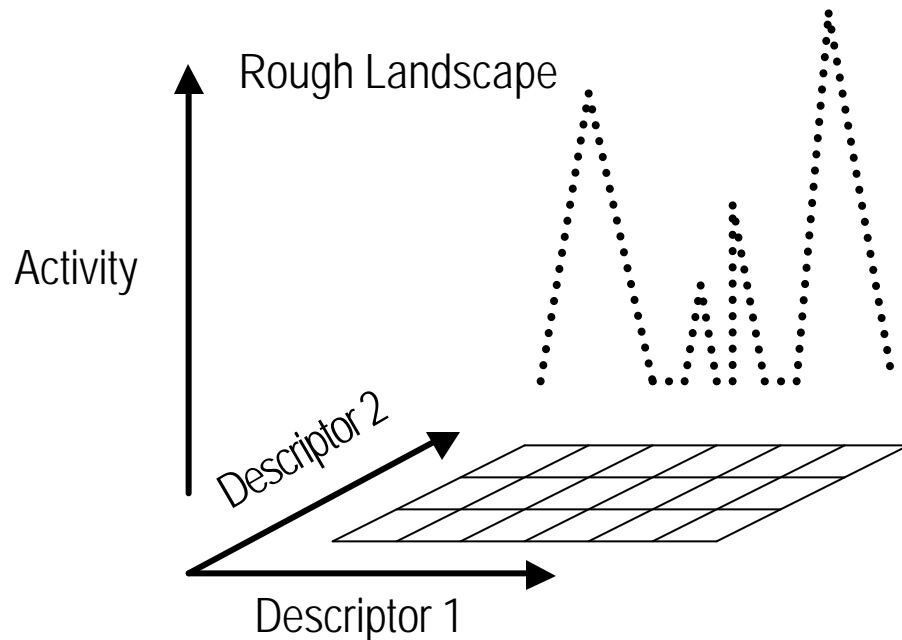
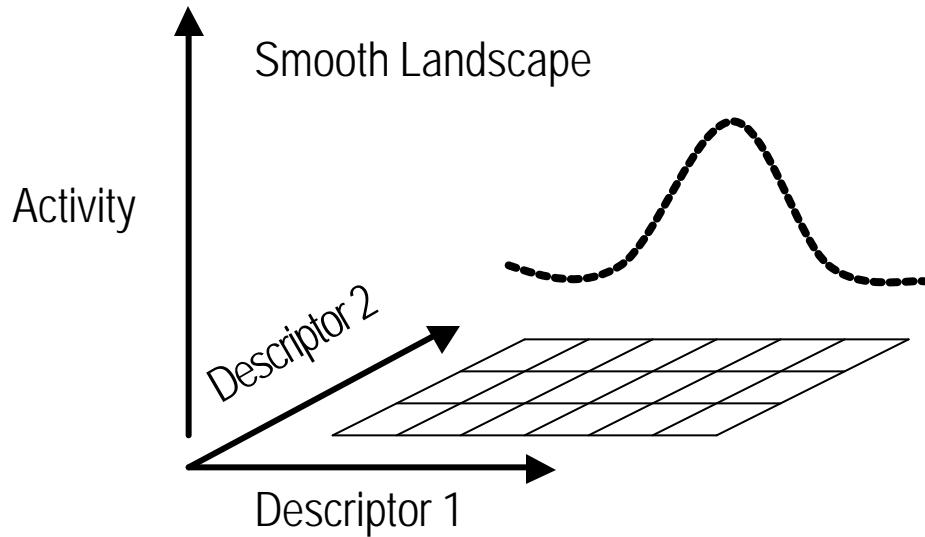
Smooth Landscape - Flint Hills, Kansas



Rough Landscape - Bryce Canyon, Utah



Why do we need to characterize them?

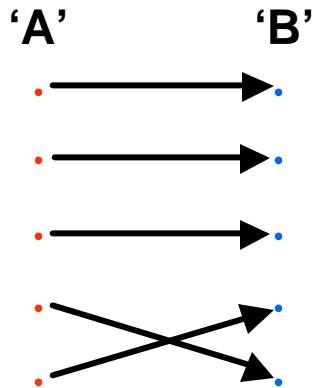


- What should be the minimum size of a representative dissimilarity subset of the corporate collection? Is it assay dependent?
- Develop “stopping-rules” to assess “have we screened enough?”
- Comparing activity landscapes of different biological targets.

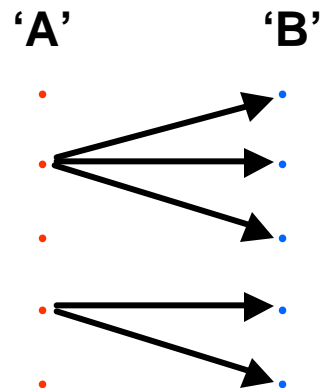
Shannon's Theory of Communication

Transmission of Messages

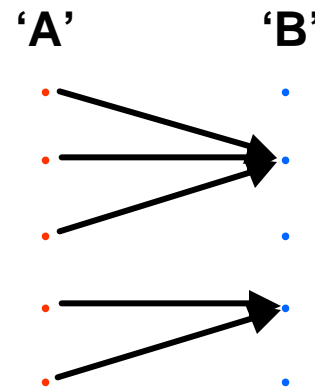
**'Perfect'
Mapping**



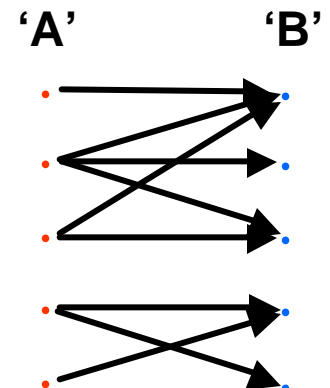
**'Noisy'
Mapping**



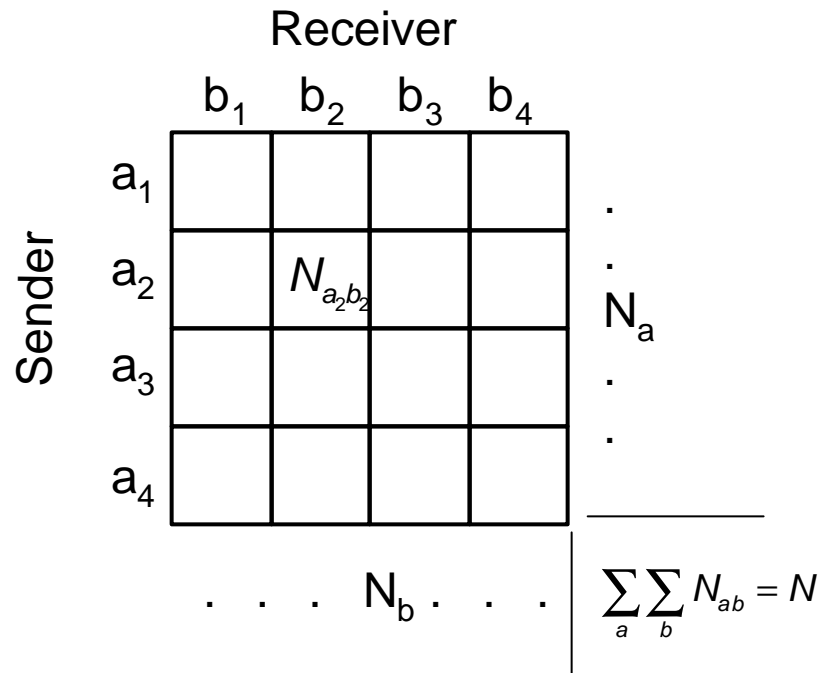
**'Equivocal'
Mapping**



**'Mixed'
Mapping**

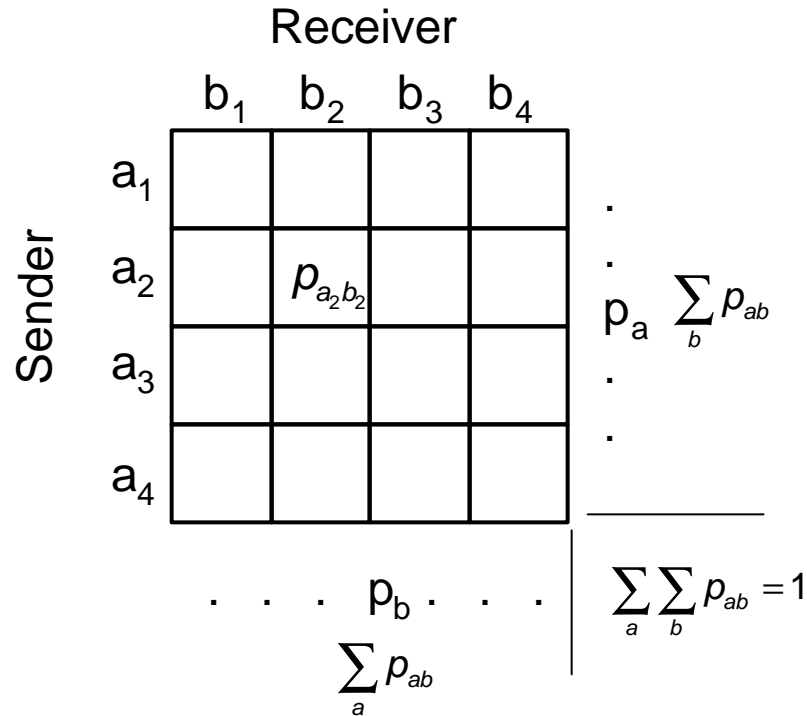


Shannon's Theory of Communication



- Probabilities or frequencies with which messages are sent.
- How each sent message is received.
- Probabilities or frequencies with which messages are received.
- How each received message was sent.

Shannon's Theory of Communication



Shannon's entropy

$$H(X) = -\sum_x p_x \log_2 p_x$$

$$H(A) = -\sum_a p_a \log_2 p_a$$

Sender's entropy

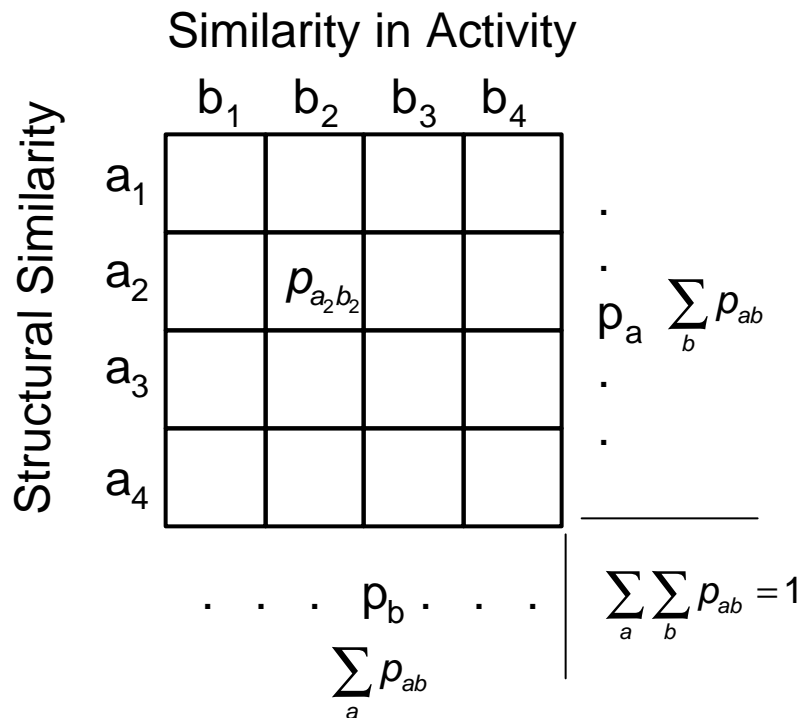
$$H(B) = -\sum_b p_b \log_2 p_b$$

Receiver's entropy

$$H(AB) = -\sum_a \sum_b p_{ab} \log_2 p_{ab}$$

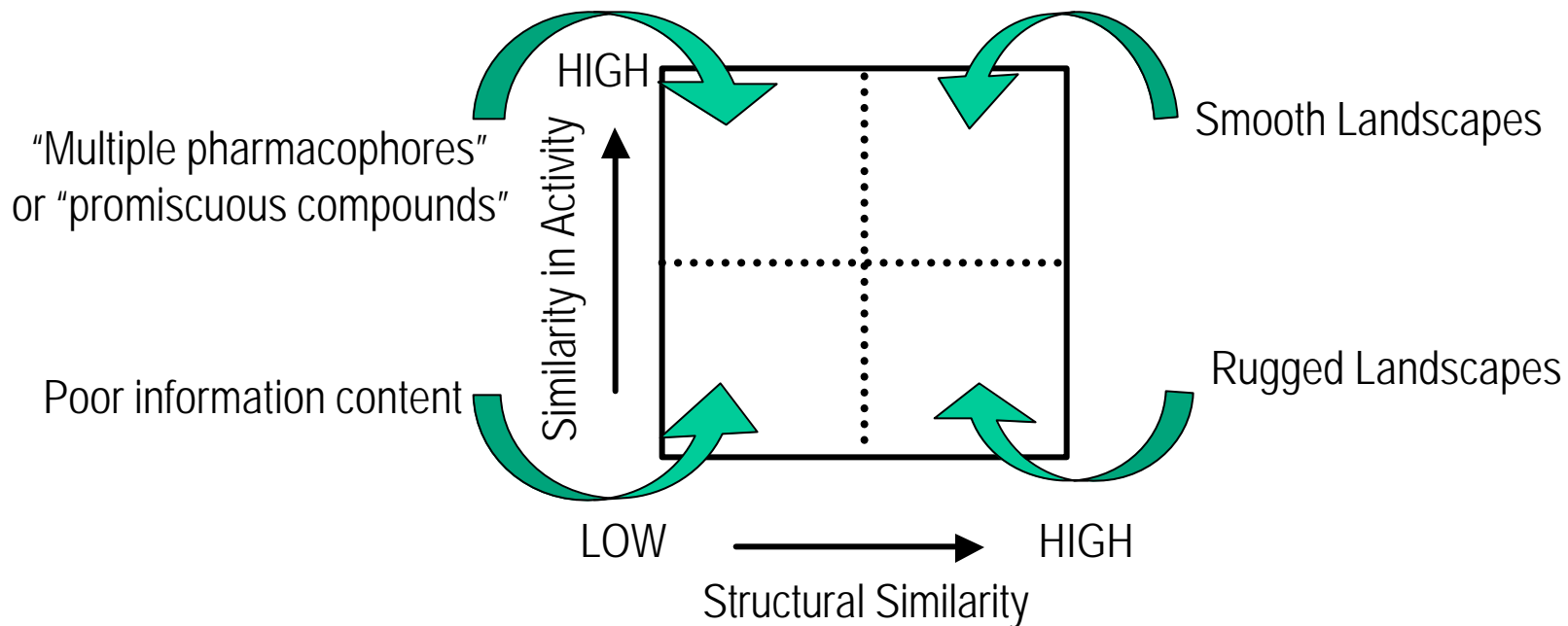
Joint entropy

Structure - Activity Mapping



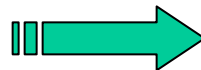
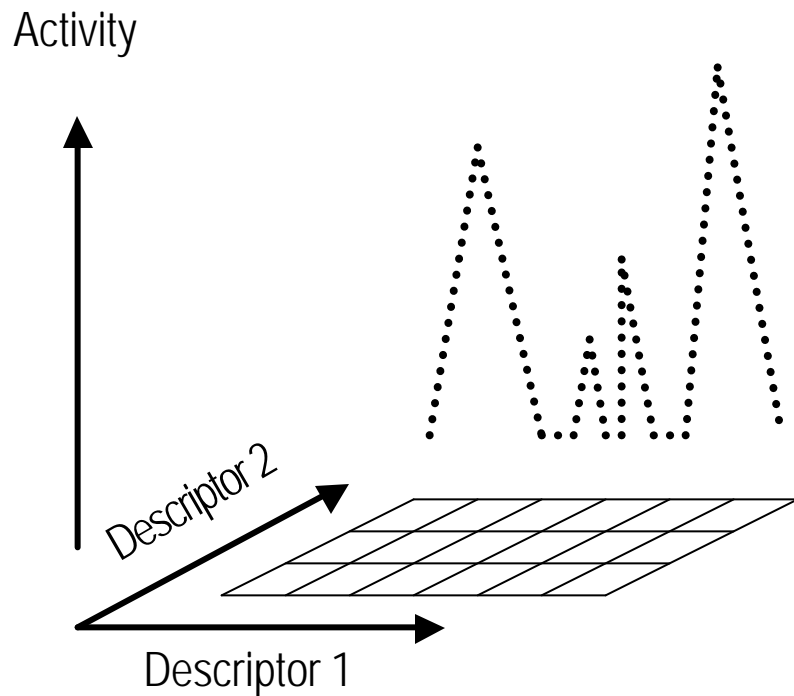
- Structural similarity - Tanimoto similarity (S_{ij}) or inter-compound distances in chemistry space.
- Activity similarity can be defined such that compounds that have similar IC_{50} or % inhibition values have a high similarity in activity.

Structure - Activity Similarity Map

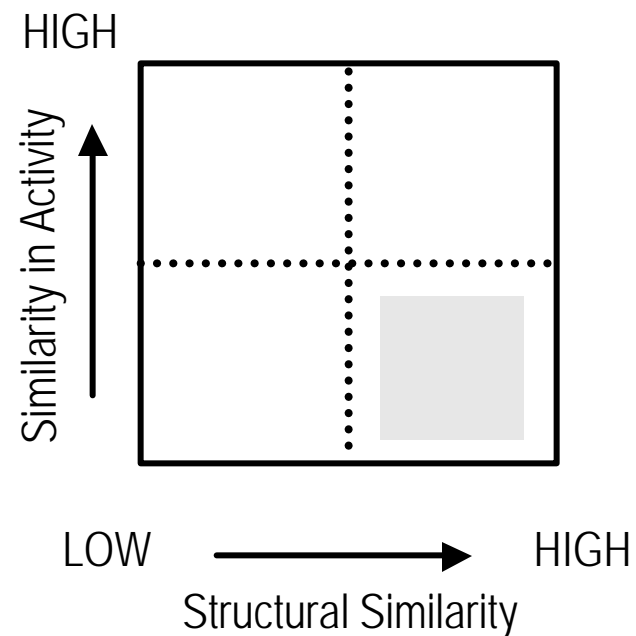


Structure - Activity Similarity Map

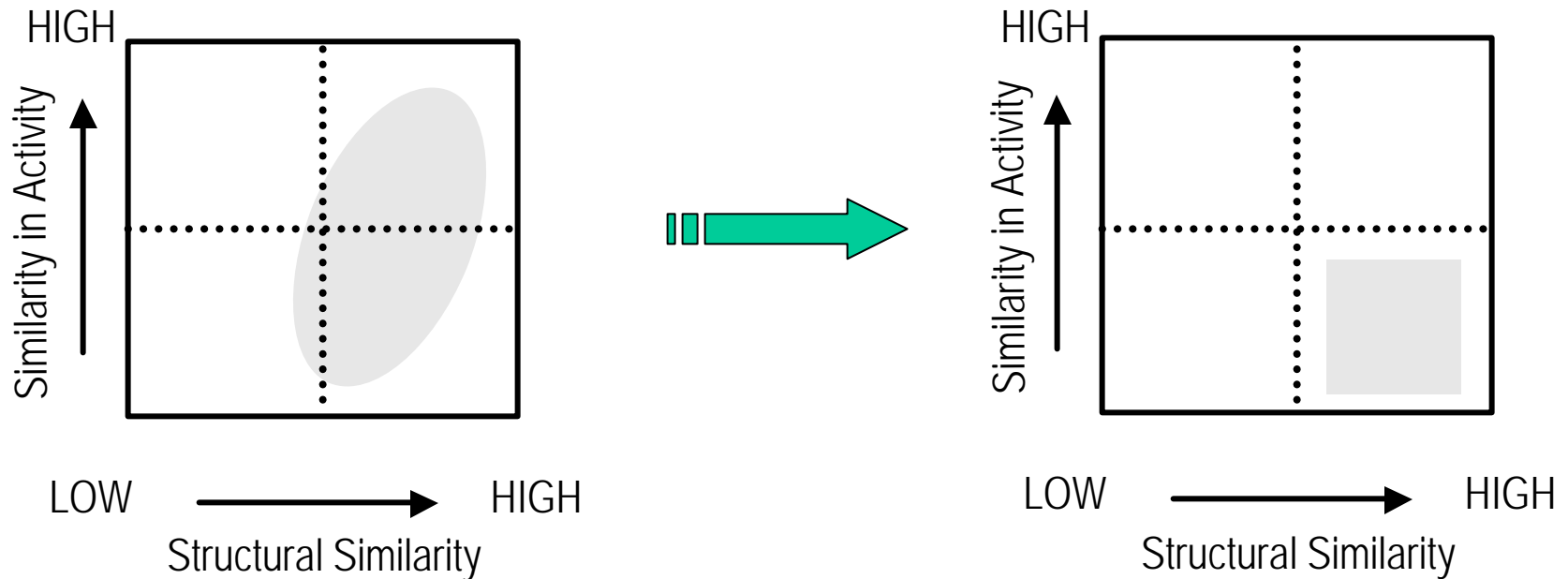
Rough Activity Landscape



Rugged Regions in Similarity Map



Information theoretic measure



Kullback-Leibler information theoretic measure could be used as a global index to characterize the topographic character of activity landscape and to compare the similarities between two different structure-activity maps.

Kullback-Leibler Index

$$D(p||q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)}$$

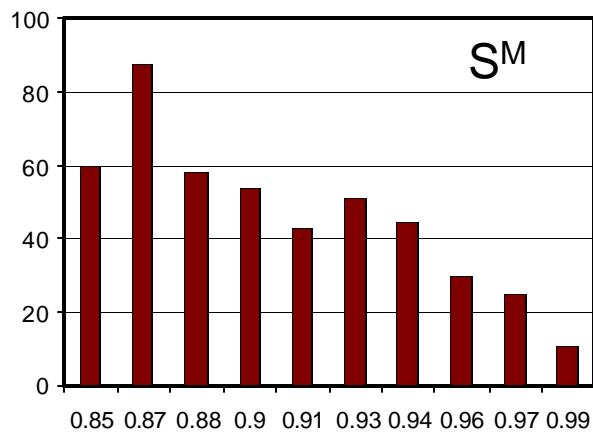
$$0 \log \frac{0}{q} = 0$$

$$p \log \frac{p}{0} = \infty$$

- Kullback-Leiber index is always non-negative.
- Index is zero, if and only if $p=q$
- Not a true "distance" - not symmetric and does not satisfy the triangle inequality.

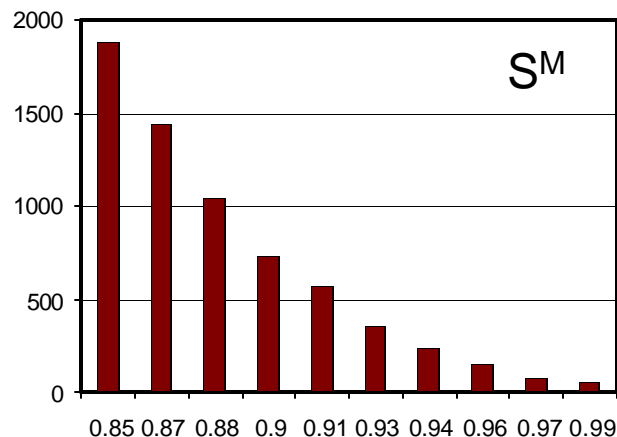
Biological Assay 1

180 cpds
16110 comparisons
465 ($S^M \geq 0.85$)



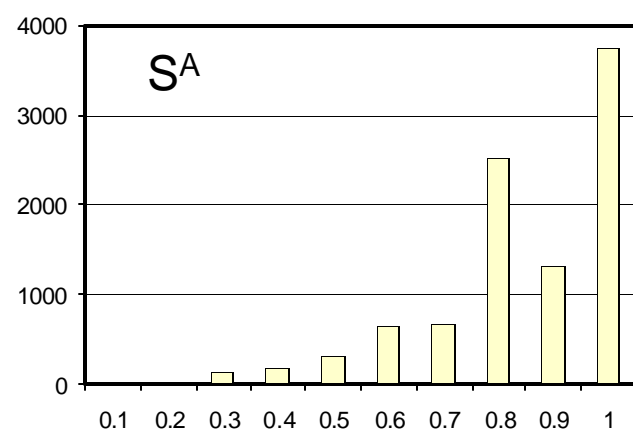
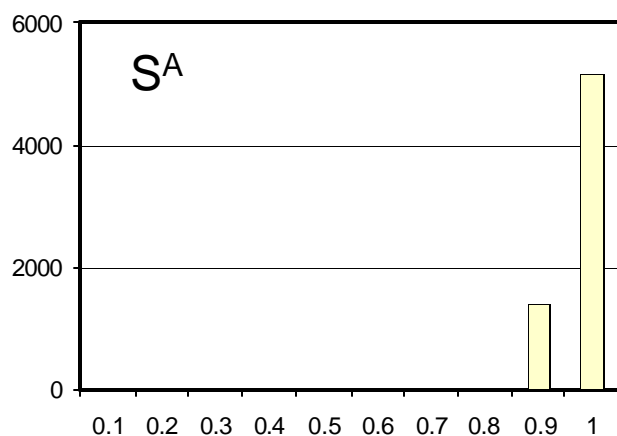
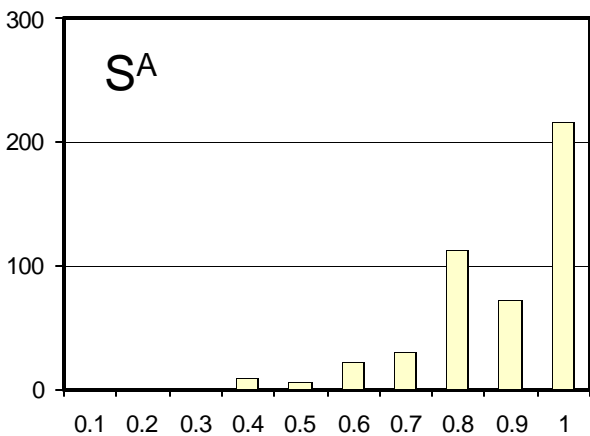
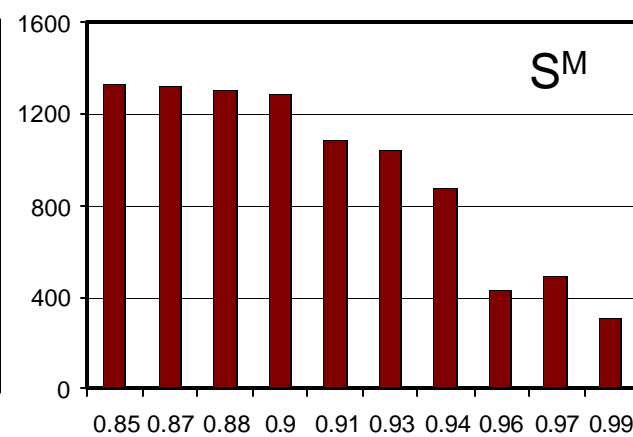
Biological Assay 2

582 cpds
169071 comparisons
6569 ($S^M \geq 0.85$)

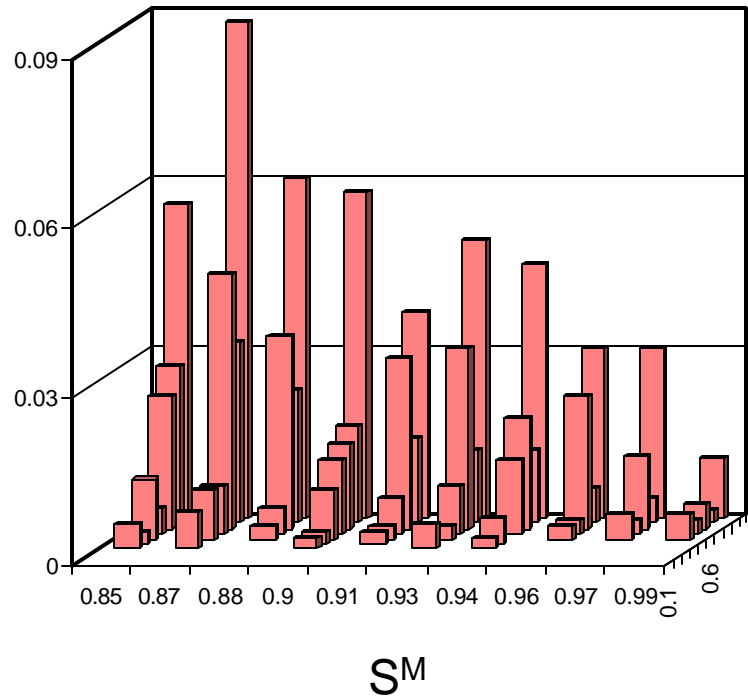
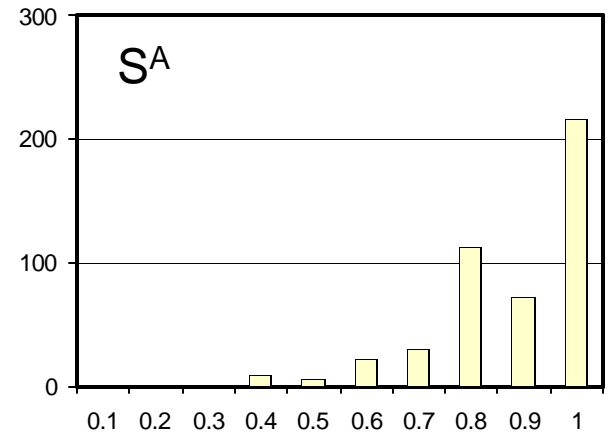
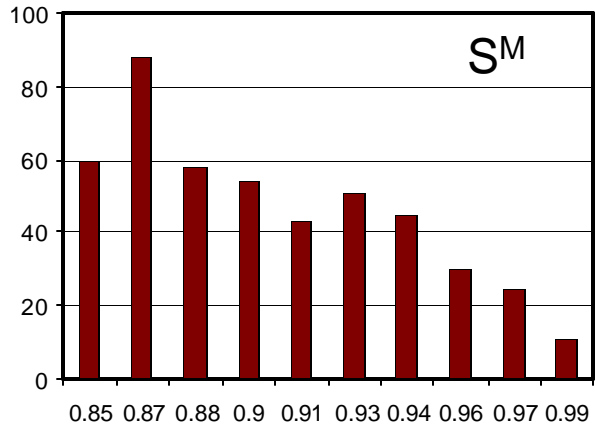


Biological Assay 3

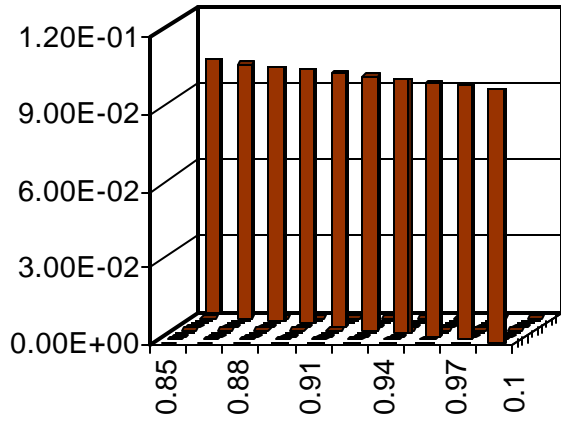
190 cpds
17955 comparisons
9508 ($S^M \geq 0.85$)



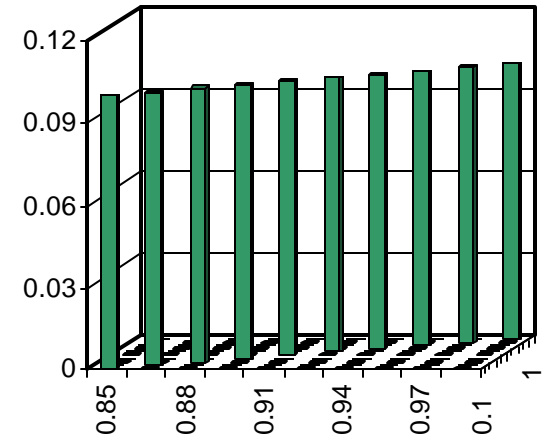
Biological Assay 1



Biological Assay 1

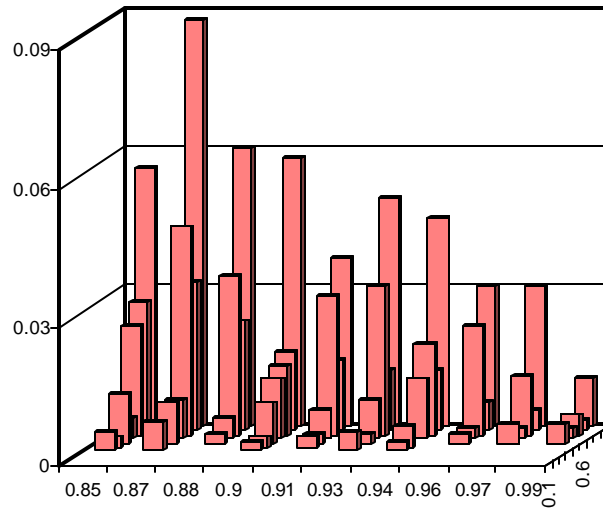


Similarity Map of an Idealized Rough Landscape



Similarity Map of an Idealized Smooth Landscape

Similarity Map of Assay 1

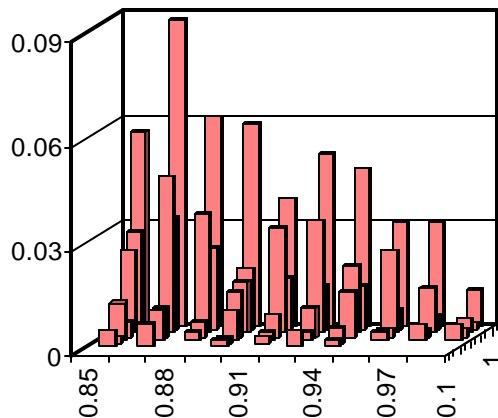


SM

SA

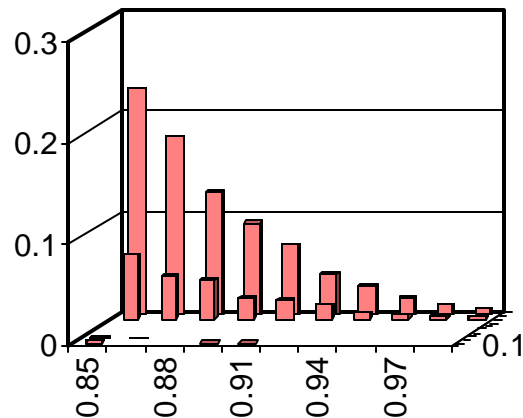
Biological Assay 1

180 cpds
16110 comparisons
465 ($S^M \geq 0.85$)



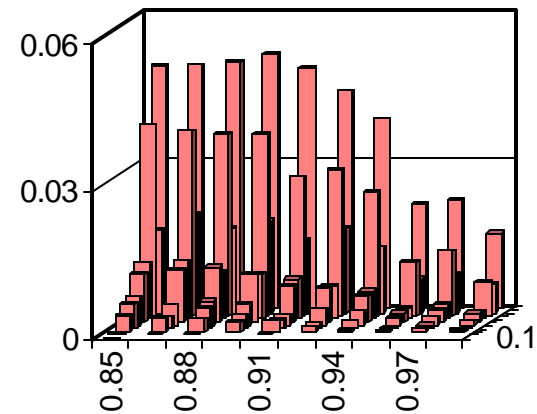
Biological Assay 2

582 cpds
169071 comparisons
6569 ($S^M \geq 0.85$)



Biological Assay 3

190 cpds
17955 comparisons
9508 ($S^M \geq 0.85$)



| DISTANCES TO IDEALIZED LANDSCAPES | ASSAY 1 | ASSAY 2 | ASSAY 3 |
|-----------------------------------|---------|---------|---------|
| SMOOTH | 9.59 | 9.39 | 9.57 |
| ROUGH | 10.68 | 12.96 | 10.36 |

Summary

- Activity landscapes tend to have smooth and rugged regions.
- Kullback-Leibler information-theoretic index can be used to measure the similarity of a given activity landscape to smooth and rough landscapes.
- If activity landscapes are like Bryce Canyon, we need to sample chemistry space more thoroughly to identify important peaks of activity.