Vol. 113, No. 1, 1983 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
May 31, 1983 Pages 178-184

THE USE OF CLUSTER ANALYSIS IN DETERMINING POSSIBLE ENZYME-SUBSTRATE INTERACTIONS WITH REFERENCE TO NUCLEOSIDE DIPHOSPHATASE ACTIVITY OF GOLGI MEMBRANES

Alison King, Alan H Fielding and Christopher A Smith*

Department of Biological Sciences
Manchester Polytechnic
Chester Street
Manchester M1 5GD, U.K.

Received March 23, 1983

The relative effectiveness of deoxyribonucleoside diphosphates, ribonucleoside diphosphates and phosphorylated B vitamins as substrates of the nucleoside diphosphatase of Golgi membranes have been determined. The substrates have been classified by cluster analysis. These findings suggest that a base ketone group and the 3'-OH of the substrates are important in promoting catalytic effectiveness, possibly by promoting hydrogen bonding during enzyme-substrate complex formation.

Enzymes are often quoted as being catalytically specific to their substrate. In general this specificity is not absolute and many enzymes show catalytic activity to a number of different substrates. Where an enzyme shows this variability, a comparison of the chemical structures of the substrates sometimes allows possible enzyme-substrate interactions to be formulated. However, this visual comparison is arbitrary and, when many substrates are involved time consuming. The present report suggests how such comparisons can be rapidly formulated by using cluster analysis to classify the substrates into groups of maximum structural similarity.

The enzyme used in this study was the nucleoside diphosphatase (E.C. 3.6.1.6) of Golgi fractions isolated from rat hepatocytes. This enzyme hydrolyses the terminal phosphate from nucleoside diphosphates;

$$NDP + H_2O \rightarrow NMP + P_i$$

*To whom all correspondence should be addressed.

Abbreviations : FMN, flavin mononucleotide; PLP, pyridoxal phosphate; PMP, pyridoxamine phosphate; TPP, thiamine pyrophosphate; PP_i, sodium pyrophosphate.

0006-291X/83 $1.50

The enzyme has been shown to be active against a variety of ribonucleoside diphosphates (1,2) and also thiamine pyrophosphate (2,3).

As part of a study of the protein components of isolated Golgi fractions we re-investigated the substrate specificity of this enzyme in Golgi membranes separated from rat liver cells. This investigation utilised a wider range of ribonucleoside diphosphate, deoxyribonucleoside diphosphates and phosphorylated B vitamins as substrates than reported in previous studies (1,2). These substrates were then classified into groups with common structural (chemical) features by cluster analysis. Chemical groups present in the most effective substrates were then identified by a novel procedure as outlined in the text. Reasons for the importance of these groups are tentatively suggested.

MATERIALS AND METHODS

All substrates were purchased from the Sigma Chemical Co., Poole, Dorset. Other chemicals were bought from accepted commercial suppliers.

Golgi fractions were isolated from rat (male, Sprague-Dawley strain) hepatocytes by the method of Morre (4) modified as described (5).

Nucleoside diphosphatase activity was measured as follows. Each assay contained 50.0 µmoles $CaCl_2$, final volume 1.0ml. Reactions were started by the addition of 100 µg Golgi protein. The assays were incubated for 30 minutes at 37°C and terminated by the addition of 1.0 ml 40% trichloroacetic acid. Precipitated protein was removed by centrifugation at 3500 g x 15 min and the phosphate content of the supernatant determined by a standard procedure (6). Appropriate control assays and standard phosphate determinations were also routinely included.

The protein content of samples was determined by the method of Lowry (7) in 0.1% sodium deoxycholate, using bovine serum albumin as standard.

Substrates were classified by cluster analysis according to the presence or absence of chemical groups eg. a ribose sugar, deoxyribose sugar, 4-keto, 5-methyl group etc. Twenty-nine possible characters were eventually assigned to the fifteen substrates analysed (Table 1). Thus each substrate was given a profile consisting of twenty-nine binary digits.

These profiles were subjected to cluster analysis using the CARM program from the PMMD package (8). This program is essentially the RELOC program from the CLUSTAN 1A package of Wishart (9). Initially each substrate is allocated to a separate group or cluster. The program then undergoes 15 cycles, during which the number of clusters or groups is progressively reduced by one, thereby producing the classification. The reduction in the number of groups is achieved by fusing (combining) the two most similar groups during each cycle. The similarities between groups were determined by calculating an error sum of squares similarity coefficient, with each of the 29 characters having an equal weighting (Table 2). A similarity coefficient of zero between two groups indicates they are identical, while progressively larger values show

increasing dissimilarity between groups. This series of fusions of increasingly dissimilar groups of substrates eventually produces the classification scheme summarised in Fig 1.

The constraints of the program will produce initially homogeneous ('real') groups but later groups may be heterogeneous ie 'artificial', because each cycle must fuse two groups. It is therefore important to recognise when heterogeneous groups are produced since this indicates that the classification is becoming increasingly arbitrary. There are two means of recognising if this stage in the fusing of groups is exceeded. The fusion of excessively dissimilar groups results in a less meaningful classification. Thus the program will switch individual substrates between existing groups to try to improve the overall classification. Secondly an error plot is produced. This shows the amount of error associated with successive fusions. If the amount of error is large the fusion has produced a heterogeneous group. Thus sudden jumps in the error plot are indicative of fusing dissimilar groups.

RESULTS AND DISCUSSION

The Golgi fractions isolated from rat hepatocytes by this method have been shown to be 70 - 80% Golgi membranes, uncontaminated by mitochondria or rough endoplasmic reticulum (10). The fraction showed a range of activity against the substrates tested (Table 1). Indeed, the fraction was active against the major deoxyribonucleoside diphosphates and phosphorylated B vitamins additional to thiamine pyrophosphate. These findings have not been previously reported.

The substrates were classified using their structural similarities by cluster analysis as described in Methods. The error plot (not shown) indicated that the substrates fall into 3 major clusters, A, B and C (Fig 1).

Groups A and B are probably 'real'. Cluster C has little internal coherence since the fusions are associated with large values of the similarity coefficient. Group C may therefore represent 3 sub-groups as indicated.

It is apparent that although groups A and B each contain a mixture of deoxy- and ribonucleotides, their average relative activities are similar ie 43 and 49% respectively. This suggests that a 3'-OH is necessary for efficient catalysis since this structural feature is common to members of both groups.

A histogram of relative activity of each substrate against its similarity coefficient to the optimum substrate (dUDP) is useful in high-lighting minor structural changes important in promoting activity. In this respect only

Table 1

Number	Substrate	Activity (%)
1	dUDP	100
2	IDP	92
3	UDP	74
4	dGDP	66
5	dTDP	33
6	GDP	31
7	CDP	21
8	dCDP	16
9	ADP	14
10	dADP	10
11	TPP	7
12	FMN	7
13	PLP	4
14	PMP	2
15	PP_i	0

Shows specificity of phosphate released by NDPase activity of Golgi membranes, as percentage of hydrolysis of dUDP. Activity was determined as described in Methods.

three pairs of substrates are important : UDP/dCDP, dGDP/TPP and IDP/ADP. This is shown in Fig 2, the remaining substrates having been omitted for clarity. It is interesting that a similar structural (chemical) difference is common to all three pairs of substrates. The active substrates, UDP, dGDP and IDP, each contain a ketone group at similar positions in their base moieties ie at the 4-pyrimidine and 6-purine positions. However, in dCDP, TPP and ADP this ketone is replaced by an amino group at the corresponding positions. Thus we suggest that a ketone group at the specified positions also promotes substrate effectiveness.

Caution must be exercised in enzymic studies using impure enzyme preparations and non-physiological substrates. However, we would tentatively conclude that the 4-keto group of pyrimidine and the 6-keto of purine

Table 2

Substrates	dUDP	IDP	UDP	dGDP	dTDP	GDP	CDP	dCDP	ADP	dADP	TPP	FMN	PLP	PMP	PP_i
dUDP	-	0.31	0.07	0.28	0.03	0.35	0.14	0.07	0.31	0.24	0.28	0.48	0.38	0.38	0.21
IDP	-		0.24	0.10	0.35	0.03	0.24	0.31	0.07	0.14	0.31	0.45	0.41	0.41	0.17
UDP	-			0.35	0.10	0.28	0.07	0.14	0.24	0.31	0.28	0.48	0.38	0.38	0.21
dGDP	-				0.31	0.07	0.34	0.28	0.17	0.10	0.34	0.48	0.45	0.45	0.21
dTDP	-					0.38	0.17	0.10	0.34	0.28	0.31	0.52	0.41	0.41	0.24
GDP	-						0.28	0.34	0.10	0.17	0.34	0.48	0.45	0.45	0.21
CDP	-							0.07	0.24	0.31	0.21	0.48	0.38	0.38	0.21
dCDP	-								0.31	0.24	0.21	0.48	0.38	0.38	0.21
ADP	-									0.07	0.31	0.45	0.41	0.41	0.17
dADP	-										0.31	0.45	0.41	0.41	0.17
TPP	-											0.34	0.31	0.31	0.21
FMN	-												0.38	0.38	0.34
PLP	-													0.07	0.31
PMP	-														0.31
PP_i	-														

Shows similarity coefficients between all possible substrate pairings. Coefficients were determined as described in Methods.

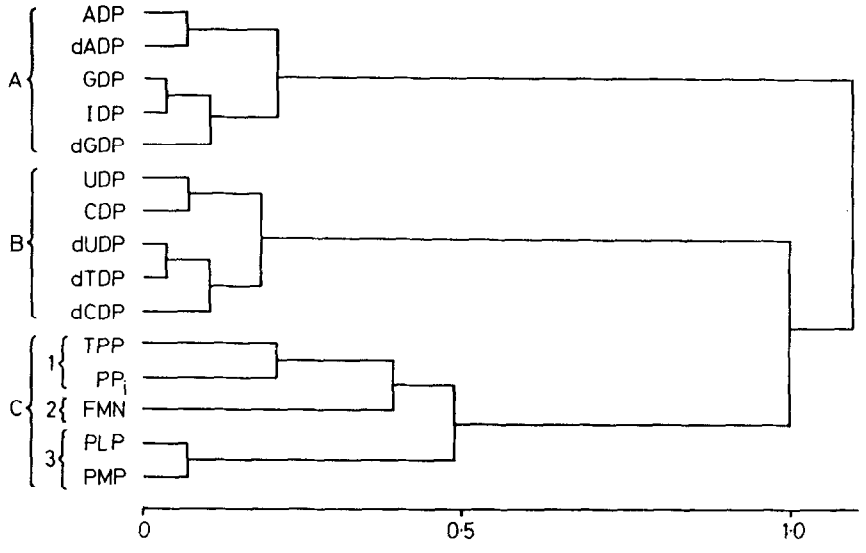

Fig 1. Dendrogram showing a classification of substrates based on their similarity coefficients as described in the text.

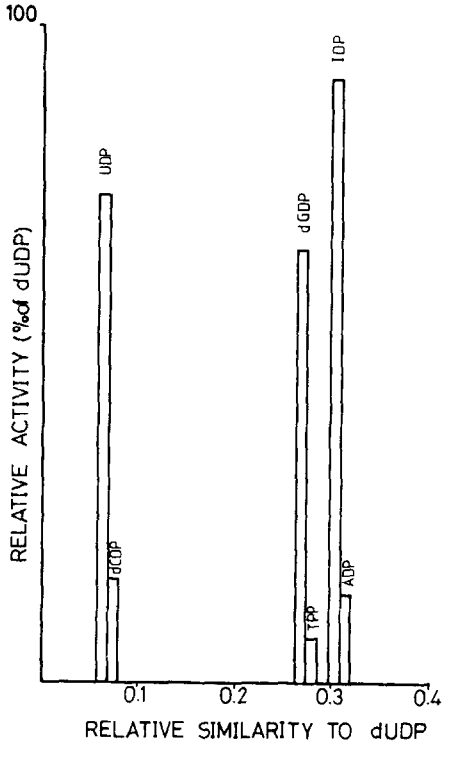

Fig 2. Histogram showing the activity of selected pairs of substrates compared to dUDP against their relative structural similarity to dUDP.

nucleoside diphosphates and the 3'-OH of the sugar moiety are necessary for optimal activity. If this is the case it is possible these groups are participating in effective hydrogen bonding during enzyme-substrate complex formation.

The above findings were reached in a rapid and easily assimilated manner by classification of the substrates by cluster analysis. This type of classification has been extensively used in the classification of a wide variety of objects (9). However, we believe this is the first time it has been applied to the investigation of enzyme-substrate interactions.

ACKNOWLEDGEMENTS

The authors would like to thank Drs D B Gordon and B Stevens-Wood for helpful discussions. This work was financed by general departmental research funds.

REFERENCES

1. Kuhn, N.J., and White, A. (1977) Biochem J. 168, 423-433.
2. Brandon, E., and Fleischer, B. (1982) Biochemistry 21, 4640-4645.
3. Yamazaki, M., and Hayashi, O. (1968) J. Biol. Chem. 243, 2934-2942.
4. Morre, D.J. (1971) in Methods in Enzymology Vol. XXII (Ed. Jackoby, W.B.) pp 130-148 Academic Press, New York and London.
5. Smith, C.A., and Brew, K. (1977) J. Biol. Chem. 252, 7294-7299.
6. Lindberg, O., and Ernster, L. (1956) in Methods of Biochemical Analysis Vol III (Ed. Glick, D.) pp 7-9 Interscience Pub. Inc., New York.
7. Lowry, O.H., Roseborough, N.J., Farr, A.L., and Randall, R.J. (1951) J. Biol. Chem. 193, 265-275.
8. Youngman, M.B. (1976) Programmed Methods for Multivariate Data Version 5, University of Nottingham School of Education, Nottingham.
9. Wishart, D. (1969) Clustan 1A:user manual. St. Andrews Computing Laboratory, University of St. Andrews, St. Andrews.
10. Smith, C.A., Powell, J.T., and Brew, K. (1975) B.B.R.C. 62, 621-626.