Prediction of the Active Sites of Proteins from Amino Acid Sequences

Naganori NUMAO* and Shun-ichi KIDOKORO

Sagami Chemical Research Center, 4-4-1 Nishi-Ohnuma, Sagami-hara, Kanagawa 229, Japan. Received April 5, 1993

The discriminant analysis of complementary units and repeated sequences of amino acids in an initial sample of 48 different enzymes produces practically useful empirical-functions which allow catalytic sites to be distinguished from non-catalytic sites. The independent variables in the discriminant functions were almost all composed of complementary units of amino acids, that is amino acid sequences whose nucleotide coding sequences were complementary to each other. In order to evaluate the validity of the functions, we applied them to the amino acid sequences of 17 different kinds of enzymes as well as 30 non-enzymes such as receptors, oncoproteins, cytokines, hormones and so on. The functions proved to be effective in predicting not only the catalytic sites of enzymes but also the binding sites of the other proteins. The result suggests that complementary units are evolutionarily conserved as a signal around the active sites of various proteins.

Keywords prediction; active site; amino acid sequence; complementary unit; discriminant analysis; hypothesis

We are interested in developing methods for the rapid prediction of the active sites of protein in the absence of knowledge of the various motifs, hydropathy, 2D or 3D structure in order to permit the rational design of molecules with agonist or antagonist functions. Mekler¹⁾ and Biro²⁾ proposed a novel hypothesis, that the specificity of interaction between amino acid residues was essentially encoded as complementary pairs of DNA (or mRNA) sequences. Blalock independently developed a similar molecular recognition theory, and showed several examples including the binding between ACTH and the peptide encoded by a nucleotide sequence complementary to the mRNA for ACTH.³⁾ However, this work is controversial; some have found it to be questionable for determination of the binding site while Fassina has claimed to have improved it.^{4,5)} The deduction of active sites in proteins by the application of such a hypothesis has not generally been demonstrated

although an important self-complementary pair in human albumin has been reported.⁶⁾ In this paper, we report a method of predicting the active sites in proteins by a statistical analysis of their sequences.

Materials and Methods

From a preliminary investigation of amino acid sequences around catalytic sites of various enzymes, we first look at repeated sequences such as $Z_1Z_2X_nZ_1Z_2$, $Z_1Z_2X_nZ_2Z_1$, $Z_1Z_1Z_1$ or $Z_1Z_2Z_1$ (Z_1 and Z_2 are an unspecified amino acid respectively, and X_n is amino acids composed of $n \leq 15$), and then at complementary units such as Z_3Z_4 and Z_3XZ_4 (Z_3 and Z_4 are amino acids with codons defined as complementary to each other, and X is an unspecified amino acid). We gave our attention to complementary units in 6 or 9 bases on DNA sequences while Biro focalized a stem and loop-structure on them. Therefore, the difference between our units and his self-complementary pairs is the distance between amino acid residues. To make a practically useful prediction for catalytic sites of enzymes, we adopted a discriminant analysis. We first constructed 128 kinds of Z_3Z_4 and Z_3XZ_4 for both directions of an amino acid (ex.

TABLE I. 48 Enzymes as Initial Sample

Enzymes	a.a.	Active sites	Enzymes	a.a.	Active sites
1. Coagulation factor IX	461	267, 315, 411	25. KEX1 carboxypeptidase	729	198, 470
2. Cathepsin G	255	62, 106, 199	26. Micrococcal nuclease	207	93, 101, 145
3. Acetylcholinesterase	649	276	27. Ribonuclease RH	238	62, 125
4. Cholinesterase	602	226	28. Cepharosporinase	381	84
5. Cutinase	228	99, 140, 208	29. ATP synthase α -chain	513	373
6. Esterase-6	548	209	30. Fructose-1,6-bisphosphatase	338	275
7. Aspartate protease	380	84, 269	31. Alkaline phosphatase	471	124
8. Rhizopuspepsin	352	60, 245	32. Fructose-bisphosphate aldorase	364	364
9. Pepsinogen	385	91, 274	33. Arginosuccinate lyase	464	51
10. Renin	410	101, 286	34. α -Enolase	434	190, 412
11. Collagenase	469	219	35. Ribulose bisphosphate carboxylase	487	204
12. Enkephalinase	749	584, 637	36. Chloramphenicol acetyltransferase	213	189
13. Cell surface protease	601	255	37. Galactose-1-phosphate uridylyltransferase	365	179, 181
14. Stomelysin	477	219	38. ADP-glucose synthase	431	39, 195
15. Thiol protease aleurain	362	168, 308	39. Creatine kinase	381	283
16. Calpain I, large subunit	714	115, 272	40. Cytochrome C peroxidase	362	116, 120, 243
17. Cysteine protease 1	343	142, 286	41. Malate dehydrogenase	312	150, 177
18. Papain	345	158, 292	42. Aldehyde dehydrogenase	497	298
19. Endoglucanase EG-1	256	149	43. Glutamate dehydrogenase	558	183
20. Lysosomal α -glucosidase	951	517	44. DNA gyrase subunit A	821	123
21. Ricin	565	201	45. Bisphosphoglycerate mutase	259	10, 61, 187
22. Shiga-like toxin I	315	189	46. DNA topoisomerase I	765	723
23. Carboxypeptidase Y	532	257, 508	47. Subtilisin BPN'	382	139, 171, 328
24. Penicillin-binding protein-6	400	66	48. Urokinase	431	224, 275, 376

a.a. is the number of amino acid. The positions of active sites are cited from their description in Swiss-prot database.

$Z_3Z_4X_nX_5$ or $X_5X_nZ_3Z_4$; X_5 is an amino acid, and X_n is amino acids composed of $n \leq 15$, and also 102 kinds of repeated sequences. We then investigated an initial sample of 48 well characterized enzymes from various sources (ex. hydrolase, lyase, transferase, oxidoreductase, isomerase and so on) providing a total of 80 catalytic sites (Table I) with the assumption that a small neighborhood around the catalytic site would affect the binding potential for substrate and catalytic rate. We defined this as all the amino acids within ± 15 from the position of the catalytic center; these were called the near while the others were called the far.

Finally, a discriminant analysis (SAS version 5; PROC Stepdisc Stepwise, $SLE=0.15$, $SLS=0.15$) of these samples was carried out to give the near function ($F_{(N)}$) and the far ($F_{(F)}$), using all the patterns to see which best discriminated the near (active site) from the far sequences, and where $F_{(N)} = C_N + \sum d_{Ni}N(Pi)$ and $F_{(F)} = C_F + \sum d_{Fi}N(Pi)$, C_N and C_F are constants associated with the real value of *a priori* probability of the near and the far amino acids in initial samples, respectively, and d_{Ni} , d_{Fi} and $N(Pi)$ are the coefficient and the number of *i* th pattern used to discriminate, respectively. The manipulation of SAS software is described elsewhere.⁷⁾ Whenever inducing a discriminant function from the enzymes, the real value of *a priori* probability of the near must be given, and 10.9% was chosen as the supposition from the ratio of the near amino acid, that is, within ± 15 of 80 active sites, to the total amino acid of the samples. Furthermore, two additional functions in the case of 30% and 50% were conveniently calculated with the same patterns, respectively. As a result of this analysis, 113 out of the total of 230 patterns could be used as independent variables of the function, and of those 93 were complementary units. Among these, the complementary units, GT, AS, GA, ID, TR, SR, LK, TXW, VXH, MXH, WXP, AXC and GXS in either direction of sequence were preferentially found in the near region. G, A, P, S, T, I,

D, R, L, K, W, V, H, M, C and X are glycine, alanine, proline, serine, threonine, isoleucine, aspartic acid, arginine, leucine, lysine, tryptophan, valine, histidine, methionine, cysteine and an unspecified amino acid, respectively. The other units were either ambiguous or located in the far region of the sequences. In addition, some non-complementary patterns were found preferentially in the far as opposed to the near.

Results and Discussion

We applied the functions to the amino acid sequences of 17 additional enzymes (Table II) to test their validity under the same conditions as those used to select the three positions for enzymes with the same number of catalytic sites as the serine proteases.⁸⁾

This showed that the method hit 16 catalytic sites, about 41% of the known peaks, and that 17 (44% of all) catalytic sites and 13 (76% of all) enzymes could be recovered. Here, a 'hit' means that the position of catalytic site is predicted within ± 15 amino acids from the known position. The result for *Bacillus stearothermophilus thermolysin* is shown in Fig. 1, while the results for all the remaining enzymes are given in Table II. We compared the prediction accuracy, that is, difference of hit efficiency between random (*ca.* 25%) and our method (*ca.* 37% on average), with that of efficiency of secondary structure prediction by three kinds of algorithms (*i.e.*, Chou, Robson, Lim; 55% on

TABLE II. The Analytical Results of 17 Enzymes

Enzymes	a.a.	Catalytic sites	Predicted positions			Hit
			10.9% ^{b)}	30% ^{b)}	50% ^{b)}	
1. Ribonuclease	124	12, 41 ^{a)} , 119	—	64, 111	29, 74, 105	2/3
2. Trypsinogen	223	40, 84, 177	53, 110	41, 103, 184	36, 94, 131, 181	2/3
3. Carboxypeptidase A	307	127, 248 ^{a)} , 270	65, 244, 291	68, 152, 238, 294	66, 115, 156, 237, 292	1/3
4. Carbonic anhydrase I	260	106, 199	56, 115	56, 119, 247	12, 58, 118, 244	1/3
5. Lactate dehydrogenase	329	99, 169, 193	103, 250	98, 148, 248, 309	30, 102, 155, 203, 247, 311	1/2
6. Superoxide dismutase	151	61, 141	—	37, 128	24, 131	1/2
7. Penicillopepsin	323	33, 213	54, 195	52, 184, 306	51, 141, 186, 227, 308	0/3
8. Triose phosphate isomerase	247	164	44, 183, 229	45, 85, 176, 227	42, 90, 169, 225	0/3
9. Alcohol dehydrogenase	374	48, 51	49, 171	47, 161, 221, 275	44, 126, 161, 220, 278, 358	1/2
10. Glutathione reductase	478	58, 63, 467	183	24, 76, 175, 358	28, 69, 105, 189, 357, 422	0/1
11. Rhodanese	293	186, 247, 249	49	56, 73, 129	19, 53, 81, 135, 191, 274	0/3
12. Catalase	506	74, 147	70	74, 266, 408, 483	14, 79, 199, 272, 416, 482	1/1
13. Lysozyme C	130	35, 53	—	27	26, 80	1/2
14. Actinidin	220	25, 162	18, 151	18, 61, 160	21, 67, 156, 188	2/3
15. Glyceraldehyde-3 phos.	333	148, 175	98, 167	26, 95, 171, 200	23, 90, 144, 169, 209, 242	1/2
16. Phospholipase A2	130	55, 106 ^{a)}	—	47, 111	42, 114	2/2
17. Thermolysin	316	143, 231	146	12, 145, 177, 225	12, 85, 141, 185, 237	1/1

a.a. is the number of amino acid. The position of catalytic sites is cited from ref. 8. a) Position is cited from the description of Swiss-prot database. b) Number indicates *a priori* probability of the near (see context). Underlined numbers indicate a position predicted within ± 15 from the catalytic site identified. A hyphen means not to be profiled.

Fig. 1. The Analytical Profiles of *B. stearothermophilus thermolysin*

The arrows indicate the position predicted as the catalytic site. The vertical and horizontal axes indicate the degree of probability of the catalytic site and the number of amino acids respectively. (—), (---) and (····) indicate the profiles when the value of *a priori* probability of the near is 10.9, 30 and 50%, respectively.

TABLE III. The Analytical Results of Non-enzymes

Proteins	a.a.	TM	Binding site ^{a)}	Predicted positions			Hit
				10.9% ^{b)}	30% ^{b)}	50% ^{b)}	
i) HIV-1, gp160	856		413—456	37, 361	30, 51, 223, 350, 530, 668, 801	27, 70, 105, 188, 229, 352, <u>469</u> , 528, 576, 676, 806	0/2
ii) CD4	458	397—418	58—83	—	248, 287, 385	38, 130, 236, 290, 387	0/5
iii) Poliovirus capsid	881		737—746	132, 200, 461, 638, 800	138, 202, 287, 385, 454, 570, 642, 798	66, 139, 195, 284, 383, 460, 573, 649, 791, 844	0/5
iv) Poliovirus receptor	417	344—367	49—123	<u>64</u>	58, 168, 271, 383	55, 163, 197, 224, 272, 338, 381	1/4
v) Insulin	110		25—54, 90—110	—	24	21	1/1
vi) Insulin receptor	1382	957—979	110—130	193, 507, 673, 762, 866, 1087	15, 191, 335, 497, 565, 692, 754, 865, 899, 902, 1022, 1087, 1167, 1232, 1346	23, <u>122</u> , 190, 244, 333, 410, 475, 507, 568, 635, 693, 753, 862, 911, 1027, 1092, 1169, 1236, 1348	0/6
vii) IL-2	153		40, 144	—	<u>128</u>	10, 100, <u>138</u>	1/3
viii) IL-2 receptor (tac)	272	242—260	22—27, 56—64	—	176	42, 143, <u>175</u> , 255	1/4
ix) TNF- α	233		77—91	<u>80</u>	88	21, 92, 159, 225	1/4
x) TNF- α receptor I	455	212—234	175—194	—	25, 50, 144, <u>193</u>	48, 152, <u>194</u> , 204, 326, 383, 441	1/4
xi) IFN- γ	166		151—154	—	38	39, <u>147</u>	1/2
xii) IFN- γ receptor	489	246—267	26—133	<u>14</u>	16, <u>76</u> , <u>128</u> , 214, 256, 332	19, <u>73</u> , <u>132</u> , 224, 261, 338, 449	3/6
xiii) Acetyl choline receptor α -subunit	461	235—259 267—285 301—320 433—451	173, 214, 216, 217	44	44, 143, <u>206</u> , 326, 394, 450	45, 103, 143, <u>207</u> , 333, 395, 444	1/6
xiv) Estrogen receptor	595		532—548	190, 463, <u>553</u>	10, 80, 189, 280, 397, 467, <u>549</u>	11, 81, 189, 246, 283, 350, 398, 465, 514, <u>548</u>	1/3
xv) A4 protein	695	625—648	597—638	141	142, 457, <u>609</u>	37, 102, 148, 194, 363, 452, 614, <u>651</u>	1/3
xvi) Glucagon	179		53—81	—	23	20, 106, 164	0/3
xvii) TGF- α	160		40—89	<u>26</u> , <u>85</u>	23, <u>79</u> , 146	22, <u>74</u> , 146	1/3
xviii) ANP	152		123—150	—	33, 75	21, <u>73</u> , <u>124</u>	1/3
xix) Angiotensinogen	484		34—43	36, 115, 264	34, 128, 186, 265, 308	33, 125, 194, 266, 335, 476	1/6
xx) Somatostatin	116		89—116	<u>104</u>	28, <u>104</u>	27, <u>104</u>	1/2
xxi) Calcitonin	136		83—114	<u>120</u>	<u>121</u>	21, <u>72</u> , <u>119</u>	1/3
xxii) Proopiomelanocortin	267		136—174, 217—224, 227—267	<u>85</u>	<u>86</u> , <u>153</u> , <u>235</u>	19, 77, <u>150</u> , <u>231</u>	2/4
xxiii) c-Ha-ras	189		12, 13, 59, 61, 63, 116, 119	<u>21</u>	<u>17</u> , <u>71</u> , <u>125</u>	<u>16</u> , <u>78</u> , <u>130</u>	3/3
xxiv) c-fos	380		143—159	108	35, 105, 302	34, 104, <u>161</u> , 225, 305, 357	1/6
xxv) p 53 suppressor gene	393		13—19, 117—142, 171—181, 234—258, 270—286	350	65, 352	34, <u>110</u> , <u>157</u> , 355	2/4
xxvi) hGH	217		28—45, 80—100, 193—217	—	135, <u>205</u>	11, 135, <u>205</u>	1/3
xxvii) NGF	307		188—195, 198	54, <u>186</u>	13, 58, 113, <u>185</u> , 252	15, 58, 120, <u>188</u> , 214, 245	1/5
xxviii) IL-6	212		186, 189, 193	—	55	12, 58, 200	1/3
xxix) HBsAg	400		21—47	151, 368	<u>31</u> , 146, 211, 364	<u>31</u> , 69, 149, 213, 291, 369	1/6
xxx) Influenza A virus	566		114, 169	23, <u>118</u> , 274, 376	22, <u>129</u> , 279, 377, 510	26, <u>117</u> , 151, 223, 280, 343, 376, 416, 511, 540	1/5

a.a. is the number of amino acid. a) The binding site is cited from reference in order, and the position reported is converted into the number of the amino acid sequence of the precursor molecule. i) L. A. Lasky, G. Nakamura, D. H. Smith, C. Fennie, C. Shimasaki, E. Patzer, P. Berman, T. Gregory, D. J. Capon, *Cell*, **50**, 975 (1987); ii) G. A. Schockmel, C. Somoza, S. J. Davis, A. F. Williams, D. Healey, *J. Exp. Med.*, **175**, 301 (1992); iii) M. G. Rossmann, E. Arnold, J. W. Erickson, E. A. Frankenberger, J. P. Griffith, H.-J. Hecht, J. E. Johnson, G. Kamer, M. Luo, A. G. Mosser, R. R. Rueckert, B. Sherry, G. Vriend, *Nature* (London), **317**, 145 (1985); iv) S. Koike, I. Ise, A. Nomoto, *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 4104 (1991); v) G. I. Bell, R. L. Pictet, W. J. Rutter, B. Cordell, E. Tischer, H. M. Goodman, *Nature* (London), **284**, 26 (1980); vi) P. De Meyts, J.-L. Gu, R. M. Shymko, B. E. Kaplan, G. I. Bell, J. Whittaker, *Mol. Endo.*, **4**, 409 (1989); vii) L. Collins, W.-H. Tsien, C. Seals, J. Hakimi, D. Weber, P. Bailon, J. Hoskings, W. C. Greene, V. Toome, G. Ju, *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 7709 (1988); viii) R. J. Robb, C. M. Rausk, M. Neeper, *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 5654 (1988); ix) S. H. Socher, M. K. Rieman, D. Martinez, A. Friedman, J. Tai, J. C. Quintero, V. Garsky, A. Oliff, *Proc. Natl. Acad. Sci. U.S.A.*, **84**, 8829 (1987); x) C. Hwang, M. Gatanaga, E. K. Inniss, T. Gatanaga, *Proc. R. Soc. Lond. B*, **245**, 115 (1991); xi) R. Wetzel, L. J. Perry, C. Veilleux, *Acad. Sci. U.S.A.*, **84**, 8829 (1987); xii) G. Garotta, L. Ozmen, M. Fountoulakis, Z. Dembic, A. P. M. van Loon, D. Stuber, *J. Biol. Chem.*, **265**, 6908 (1990); G. Chang, *Protein Engineering*, **3**, 611 (1990); xiii) G. Garotta, L. Ozmen, M. Fountoulakis, Z. Dembic, A. P. M. van Loon, D. Stuber, *J. Biol. Chem.*, **265**, 6908 (1990); xiv) T. M. Dennis, J. Giraudat, F. Kotzyba-Hilbert, M. Goeldner, C. Hirth, J.-Y. Chang, C. Lazure, M. Chretien, J.-P. Changeux, *Biochemistry*, **27**, 2346 (1988); xv) A. Ratajczak, S. P. Wilkinson, M. J. Brockway, R. Hahnel, R. L. Moritz, G. S. Begg, R. J. Simpson, *J. Biol. Chem.*, **264**, 13453 (1989); xvi) J. Kang, H.-G. Lemaire, A. Unterbeck, J. M. Salbaum, C. L. Masters, K.-H. Grzeschik, G. Multhaup, K. Beyreuther, B. Muller-Hill, *Nature* (London), **325**, 733 (1987); xvii) G. I. Bell, R. Sanchez-Pescador, P. J. Laybourn, R. C. Najarian, *Nature* (London), **304**, 368 (1983); xviii) R. Derynck, A. B. Roberts, M. E. Winkler, E. Y. Chen, D. V. Goeddel, *Cell*, **38**, 287 (1984); xix) R. Kageyama, H. Ohkubo, S. Nakanishi, *Biochemistry*, **23**, 3603 (1984); xx) L.-P. Shen, R. L. Pictet, W. J. Rutter, *Proc. Natl. Acad. Sci. U.S.A.*, **79**, 4575 (1982); xxi) E. Poeschl, I. Lindley, E. Hofer, J. M. Seifert, W. Brunowsky, J. Besemer, *FEBS Lett.*, **226**, 96 (1987); xxii) G. Boileau, C. Barbeau, L. Jeannotte, M. Chretien, J. Drouin, *Nucleic Acids Res.*, **11**, 8063 (1983); xxiii) F. Jurnak, *TIBS*, **13**, 195 (1988); xxiv) R. Turner, R. Tjian, *Science*, **243**, 1689 (1989); xxv) T. Soussi, C. C. de Fromental, P. May, *Oncogene*, **5**, 945 (1990); xxvi) B. C. Cunningham, J. A. Wells, *Science*, **244**, 1081 (1989); xxvii) P. Cohen, A. Sutter, G. Landreth, A. Zimmermann, E. M. Shooter, *J. Biol. Chem.*, **255**, 2949 (1980); xxviii) C. Nishimura, K. Fukunaga, K. Yasukawa, T. Kishimoto, Y. Arata, *FEBS Lett.*, **281**, 167 (1991); xxix) A. R. Neurath, S. B. H. Kent, N. Strick, K. Parker, *Cell*, **46**, 429 (1986); xxx) W. Weis, J. H. Brown, J. C. Paulson, J. J. Skehel, D. C. Wiley, *Nature* (London), **333**, 426 (1988). b) Number indicates *a priori* probability of the near (see context). A hyphen is the same as in Table II. TM is transmembrane region. Underlined numbers indicate a position predicted within ± 15 from the binding site reported.

average) and random (ca. 38%).⁹⁾ The prediction accuracy of the two procedures is the almost same, that is, the 'hit' efficiency is ca. 1.5 times more effective than the random prediction. As might be expected from the codon symmetry of complementary units in the patterns, similar active sites were found when the functions were applied to the amino acid sequence encoded by the complementary nucleotide sequence to the mRNA of a given protein (data not shown). We noticed that the functions could be applied to non-enzymes because they successfully predicted the binding site of tissue plasminogen activator (t-PA) to endothelial cell in addition to predicting the positions of the sites involved in catalysis.¹⁰⁾

Next, we applied the functions to 30 different non-enzymes including receptors, onco-proteins, cytokines and hormones (Table III). Surprisingly, one of the predicted positions was shown to have good fit to at least one of the reported binding site (Table III). In the case of those proteins whose active forms were composed of less than 50 amino acids such as insulin, glucagon, A4 protein, TGF- α , ANP, angiotensin I, somatostatin, calcitonin or ACTH, the biologically active regions hidden in the longer precursor molecules could be found with good efficiency. In total, hits were 31 (ca. 26%) of the predicted positions and 25 (ca. 83%) of all samples examined could be recovered with the conditions that a maximum of six positions could be chosen for each protein. In this case, a 'hit' means that a part or all of a binding site is contained within ± 15 from the predicted position. However, the method did not predict a site in CD4 or poliovirus. In the case of glucagon, although the units SXG and GT exist at positions 54 to 57 in the amino acid sequence, a shoulder can be found because of the existence of other units SR and AS in the signal region.

The number of test samples may still not be sufficient to discuss the generality between catalytic and binding sites. However, the 'hit' efficiency in non-enzymes could be increased to about 28% if the positions predicted in such a signal or/and cytoplasmic region could be excluded in advance. Furthermore, the fact that functions based on the catalytic sites of the enzymes could give roughly the same efficiency in the prediction of binding sites may indicate the possibility that complementary units or something

like Mekler's or Biro's concept is conserved as a signal around both active sites. More interestingly, the active sites predicted by our discriminant functions partially overlapped with that of Gabrielian's criteria¹¹⁾ seeking a minimum value of Z_1XZ_2 (Z_1 , Z_2 and X are an unspecified amino acid, respectively) using the frequency of occurrences of amino acids, which could predict the active site of proteins composed of less than 100 amino acids. Gabrielian's criteria, however, are not effective in longer protein sequences (data not shown). Taking the relation into consideration, therefore, we are now in the process of simplifying complementary units to construct a set of universal motifs and to solve the problem in Blalock's procedure³⁾ from a new point of view. Our present functions are quite effective in finding unknown functional sites in protein sequences themselves. We (N. N.) have recently demonstrated another binding site of tumor necrosis factor receptor I to TNF- α as shown in Table III to rationally design molecules with an antagonist function.¹²⁾

Acknowledgements We thank Sagami Chemical Research Center, Nippon Mining Co., Ltd. and Tosoh Corporation for their financial support of this research, and Drs Y. Ueda, H. Asaine, M. Sano and Y. Eguchi at MKI for help in programming and useful suggestions. We appreciate very much Dr. S. Brenner at MRC and Dr. K. Kondo, President of SCRC for critical comments on the manuscript.

References

- 1) L. B. Mekler, *Biofizika*, **14**, 581 (1969).
- 2) J.-C. Biro, *Medical Hypotheses*, **7**, 981 (1981).
- 3) J. E. Blalock, *TIBTECH*, **8**, 140 (1990).
- 4) A. N. Eberle, R. Drozd, J. B. Bauman, J. Girard, *Peptide Res.*, **2**, 213 (1989).
- 5) G. Fassina, G. Cassini, *Biochem. J.*, **282**, 773 (1992).
- 6) K. G. Draper, *Biochem. Biophys. Res. Commun.*, **163**, 466 (1989).
- 7) "SAS User's Guide Statistics," Version 5 ed., ed. by SAS Institute Inc., Tokyo, 1985.
- 8) M. J. J. M. Zvelevil, M. J. E. Sternberg, *Protein Engineering*, **2**, 127 (1988).
- 9) W. Kabsch, C. Sander, *FEBS Lett.*, **153**, 179 (1983).
- 10) A. J. Wittwer, M. A. Sanzo, *Thromb. Haemostas.*, **64**, 270 (1990).
- 11) A. E. Gabrielian, V. S. Ivanov, A. T. Kozhich, *CABIOS*, **6**, 1 (1990).
- 12) B.-L. Lie, D. Tunemoto, H. Hemmi, Y. Mizukami, H. Fukuda, H. Kikuchi, S. Kato, N. Numao, *Biochem. Biophys. Res. Commun.*, **188**, 503 (1992).