TASK QUARTERLY 6 No 3 (2002), 469–482

# CORRELATED MUTATIONS IN SELECTED PROTEIN FAMILIES

JACEK LELUK[1,2], MONIKA SOBCZYK[1]
AND ŁUKASZ BECELLA[1]

[1] *Institute of Biochemistry and Molecular Biology, University of Wroclaw, Tamka 2, 50-137 Wroclaw, Poland*

[2] *Interdisciplinary Centre for Mathematical and Computational Modeling (ICM), University of Warsaw, Pawińskiego 5A, 02-106 Warsaw, Poland*
*lulu@bf.uni.wroc.pl*

(Received 30 September 2001; revised manuscript received 31 January 2002)

**Abstract:** Four different protein families (two proteinase inhibitor families, myoglobins and lysozymes) were surveyed for correlated mutations with respect to the position distance and their significance in structure stabilization and biological activity. They were chosen for this study in order to verify the currently admitted model of mutational correlation relationship with respect to spatial contact of the residues and contribution in protein biological activity. There was observed high contribution of spatially dispersed residues (which are also not involved in the protein active center) in mutational correlation. Because of the significantly large distance between correlated positions these cases do not correspond explicitly to any mechanism included in current hypotheses. It is suggested that the role of residue spatial contact in structure preservation, intermolecular interaction and active site rescue mechanisms only partially explains the correlation phenomenon.

**Keywords:** protein sequence, multiple alignment, tertiary structure, mutational correlation, genetic semihomology algorithm

## 1. Introduction

The neutral theory of molecular evolution [1] proclaims that most fixed mutations are selectively neutral. The negative Darwinian selection usually serves as the deletion process that rejects the lethal or unprofitable changes. The positive selection is considered as rarely occurring. However, there was described the importance of positive Darwinian selection among protein inhibitors of serine proteinase [2, 3]. It was stated that the positive selection concerns especially the molecule fragments responsible for their biological activity (antiproteinase reactive site). Moreover, the positive selection at the reactive sites undergoes more intensive accelerated evolution than at the other fragments of inhibitor molecule [4].

According to the currently assumed model the positive mutations do not occur independently. They are related to the changes occurring in their neighborhood, they

reflect the protein-protein interaction and they must preserve the biological activity and structural properties of the molecule. Therefore the fixed mutations should be associated with other fixed mutations that occur simultaneously. The phenomenon of several mutations occurring simultaneously is known as correlated mutations. There are many reports that confirm the relationship between correlated mutations and direct contact of involved residues [5, 6]. The correlation concerns the conservation of the local charge and/or the amino acid side chain metric volume. The basic purpose of correlated mutations in this case is to preserve the structural consensus characteristics. The correlated changes reported for myoglobin family deal with the residues that are in contact in tertiary structure [6].

Another kind of correlation concerns the protein-protein interaction. According to the reported model [7], a mutation in one of the interacting protein "forces" appropriate change in the other protein, so that to optimize the interaction. It is rather impossible to estimate which mutation is preceding, the changes usually must appear simultaneously, otherwise each of them is rejected as unfavorable or lethal. The mutations concerned to protein-protein interaction do not refer to the contact within one molecule, but to the residues involved in intermolecular interaction, therefore they usually are assembled into limited clusters on one side of the protein surface.

In this paper four different protein families were surveyed for correlated mutations with respect to the position distance and their significance in structure stabilization and biological activity. The examined groups concerned two proteinase inhibitor families, myoglobins and lysozymes. The Bowman-Birk inhibitor family is characterised by double-headed nature (structural as well as functional) and very high contents of cysteine (14) that form 7 disulfide bridges. The eglin-like proteins are proteinase inhibitors of different stabilization properties (lack of cysteines). Two other families (myoglobins and lysozymes) are among the best-described groups according to their structure and function. They were chosen for this study in order to verify the currently admitted model of mutational correlation relationship with respect to spatial contact of the residues and contribution in protein biological activity.

## 2. Materials and methods

The protein sequences were obtained from three databases: SWISS-PROT, TREMBL (http://www.expasy.ch) [8–12] and NCBI (http://www.ncbi.nlm.nih.gov).

The preliminary homology search was performed with the use of BLAST search tools (http://www.ncbi.nlm.nih.gov/BLAST) [13] and homology degree verified with the genetic semihomology algorithm [14–16].

The sequence multiple alignment was constructed with the use of genetic semihomology algorithm [14–16] and program SEMIHOM.

The correlated mutation search was carried out with the use of simple DOS program FEEDBACK (C++ compiled) yielding the data that are further processed with Microsoft Excel application for final visualization. The program FEEDBACK returns in result all possible residues occurring at all sequence positions of a protein family for each residue occurring at each position.

The protein tertiary structure and the location of correlated positions were analyzed with Insight II (Molecular Simulations Inc.) and visualized with the WebLab® Viewer (Molecular Simulations Inc.).

The sequence numbering considers the gaps inserted during multiple alignment. The gaps were taken into account in the FEEDBACK tables. Therefore the residue numbers may differ from those shown on molecule tertiary models (with the larger numbers the more is the difference). That is because the gaps cannot be expressed on the tertiary structures.

The programs SEMIHOM and FEEDBACK are freely available upon request to the authors of this article (lulu@bf.uni.wroc.pl).

## 3. Results and discussion

Table 1 demonstrates the number of observed correlation sets and their surface distribution characteristics in four studied protein families. The correlation sets are classified to three general subgroups. The most interesting subgroup concerns the dispersed correlations. They refer to the positions correlated mutationally and very distant from each other. Those positions are not in direct contact, they also cannot reveal any interaction. For example the distance between alpha carbons of correlated residues is up to over 22 Å for eglins, 18–19.5 Å for Bowman-Birk inhibitors, 14–18.5 Å for myoglobins and 12–24.3 Å for lysozymes. There are no other correlated changes observed between those amino acids. The other two subgroups include correlated positions that are in spatial contact with each other or they are located close enough to interact. Depending on the distribution of these residues the latter two subgroups form narrow clusters, or undirected spots. In some cases it is hard to classify the correlation sets to a certain subgroup. Some sets may consist of both, closely located residues and distant residues, and all are mutationally correlated. Therefore the results presented in Table 1 should be considered as the general outline rather than the precise result. The correlation sets of mixed character are classified to the selected subgroup depending on the distribution characteristics of most of the involved residues.

The contribution of relatively distant residues in mutational correlation is very high or significant, except for one family (Bowman-Birk inhibitors). It often concerns the positions scattered over entire surface of a molecule. The residues involved in these sets in many cases have no relationship to the regions directly responsible for the biological activity.

The selected eglin-like proteins include 25 sequences of the highest similarity score to eglin C from *Hirudo medicinalis* according to the algorithm of genetic semihomology. The sequence multiple alignment reveals 47 consensus positions out of 70 positions aligned (including gaps) (Figure 1). There were found and analyzed 20 sets of the mutationally associated variable positions. The sets included 2 to 13 positions. Among them 7 sets contained distantly scattered residues and 13 groups of residues located close enough to form a cluster at a certain part of the molecule surface. Seven clusters were narrow in shape, and formed a kind of correlation path. Six other clusters formed unshaped spots. The typical sets of each group are presented in Figure 2. Among 20 observed sets only one set contained the residue involved in the reactive site formation.

**Table 1.** The observed number and contribution of three correlation types in four different protein families. The correlation sets consist of 2 to over 20 residues

| The protein family (number of correlated positions/set) | The correlation statistics | | | | |
|---|---|---|---|---|---|
| | Total number of correlation sets observed | Number of dispersed correlations | Number of narrow clusters | Number of undirected clusters | Number of correlations related to active center |
| Eglin-like proteins (2–13) | 20 | 7 | 7 | 6 | 1 |
| Bowman-Birk proteinase inhibitors (2–28) | 23 | 4 | 13 | 6 | 9 |
| Myoglobins (2–29) | 41 | 23 | 9 | 9 | n. a. |
| Lysozymes (2–15) | 41 | 25 | 9 | 7 | 9 |
| All families | 125 (100%) | 59 (47.2%) | 38 (30.4%) | 28 (22.4%) | – |

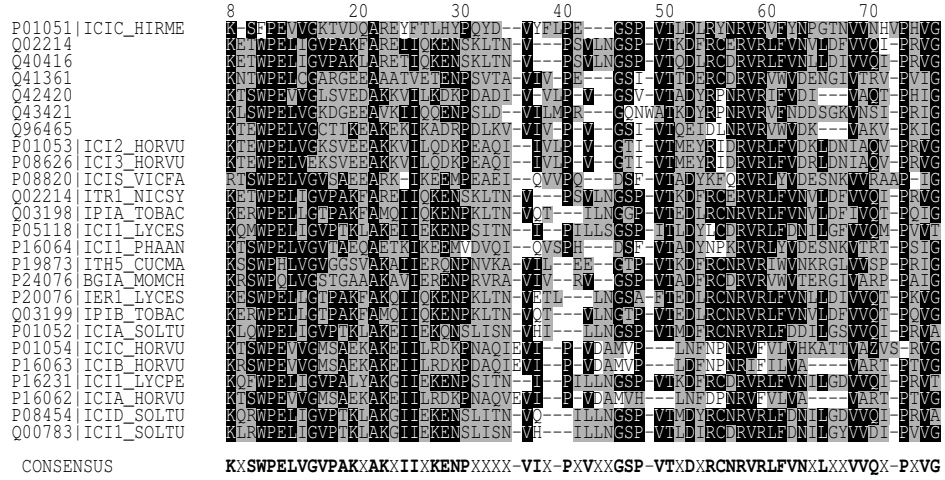



**Figure 1.** The multiple alignment of proteins homologous to eglin C from *Hirudo medicinalis* (P01051), constructed according to genetic semihomology algorithm; the conservative consensus residues are shown as white letters on black background; the gray background indicates genetic semihomology between aligned residues

The study of Bowman-Birk inhibitor family (52 selected sequences) revealed 23 sets of correlated positions. These sequences are highly homologous to each other. The multiple alignment (Figure 3) verified by genetic semihomology algorithm disclosed 52 consensus positions out of 64 compared. Among the mutationally correlated sets four refer to dispersed positions, 13 form narrow clusters and 6 are gathered into wide-shaped spot clusters. Fourteen sets include only residues that are not involved in the reactive site(s) formation. The particular sets consist of 2 to 28 positions. The typical sets of different character (dispersed and clustered) are shown in Figure 4.

The eglin-like proteins and Bowman-Birk inhibitor family act as the proteinase inhibitors. Although they have similar biological function, the structure properties and structure stabilization mechanisms are different. It is reflected by the difference in amino acid composition of these two families (Figure 5). The structure of Bowman-Birk inhibitors is stabilized mainly by large amount of disulfide bridges, which are

(a)

| Position number and occurring residues | Correlation versus position 67 | |
|---|---|---|
| **67 [-DGNT]** | **D (8)** | **G (9)** |
| 10 [-ELNQRST] | ET | LNQRS |

(b)

| Position number and occurring residues | Correlation versus position 34 | |
|---|---|---|
| **34 [DEKQRST]** | **T (10)** | **Q (6)** |
| 15 [CILV] | CIL | V |
| 17 [DGPRST] | PR | ST |
| 27 [EKLQT] | EQ | KL |
| 28 [AEKLQRT] | KT | EQR |
| 30 [KMNRY] | N | KM |
| 32 [DEKLNQRS] | KLS | DEN |
| 56 [CFILPY] | C | IP |
| 68 [-DFIKLNSTY] | DFI | -KNT |

(c)

| Position number and occurring residues | Correlation versus position 30 | |
|---|---|---|
| **30 [KMNRY]** | **K (6)** | **N (15)** |
| 18 [DGPRST] | S | DGPRT |
| 27 [EKLQT] | L | EQ |
| 29 [DEHQ] | D | EQ |
| 33 [AILVY] | A | ILV |
| 35 [-AINV] | IV | -AN |
| 68 [-DFIKLNSTY] | -NS | DFIKLSY |

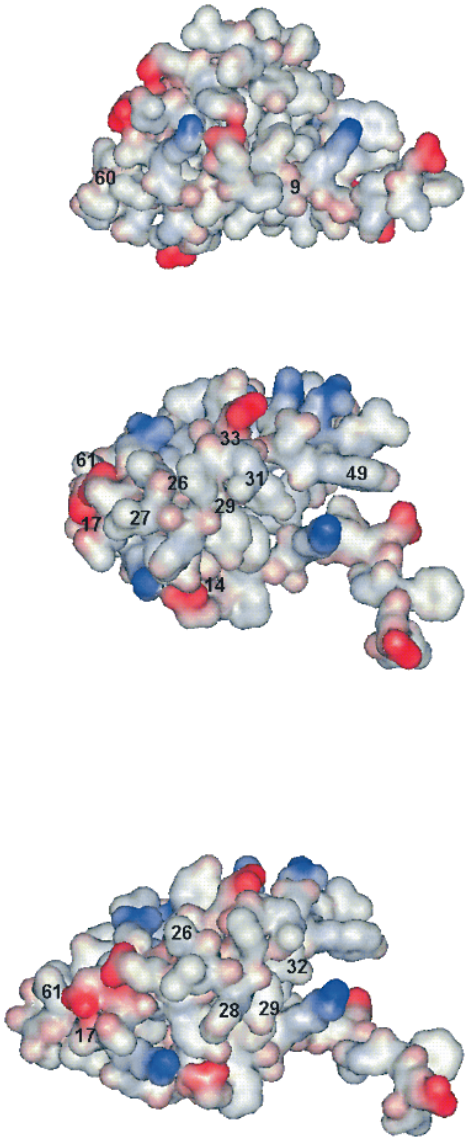



**Figure 2.** The three types of distribution of correlated positions present in eglin-like proteins: (a) the dispersed correlation, (b) the narrow correlation cluster, (c) the spot correlation cluster; the residue location and relative distribution is shown on tertiary structure of eglin C (P01051)

absent in eglins. Therefore the cysteine contribution in Bowman-Birk amino acid composition is very high, and in eglins it is extremely low. The high contribution of disulfide bridges makes the structure rigid and difficult to deform. For that reason the variability of the other positions of Bowman-Birk inhibitors can be unconstraint without affecting the principal structure properties. The most variable positions are occupied by up to 12 residues of very different nature (Figure 6). The structure of eglin-like proteins is stabilized mainly by hydrophobic residues whose relative

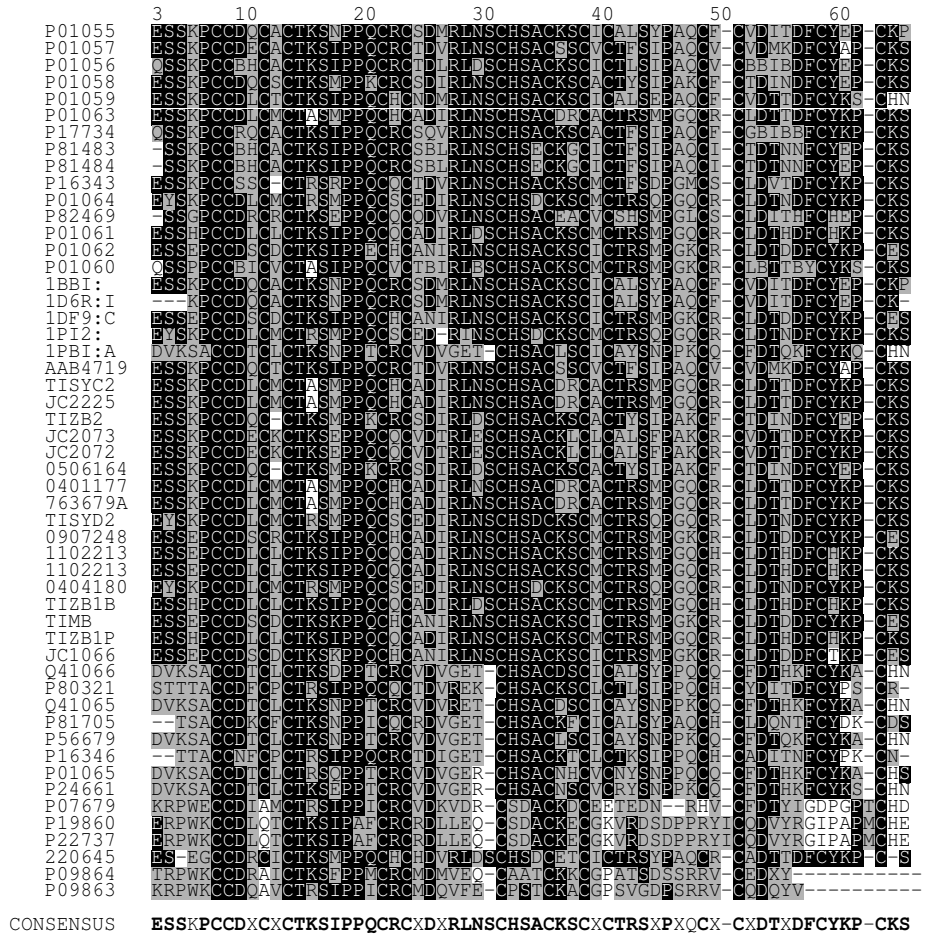



**Figure 3.** The multiple alignment of Bowman-Birk proteinase inhibitors, constructed according to genetic semihomology algorithm; the conservative consensus residues are shown as white letters on black background; the gray background indicates genetic semihomology between aligned residues

amount is higher comparing to the Bowman-Birk inhibitors. Although the variability in eglin family is also high, the maximum variability does not exceed 9 residues (Figure 6). Despite the mechanism of structure determination and stabilization is different, the general outline of mutational correlation is similar in both cases. That would suggest that the correlation assignment does not concern mainly structure preservation. The low contribution of the positions involved into reactive site formation denies the coherence between mutational correlation and biological activity establishment.

The multiple alignment of 74 myoglobins shows high homology within this family (Figure 7). Every sequence position has a significant meaning for consensus outline, *i.e.* there can be found a typical consensus residue for each position of the sequence. The correlated positions are grouped into sets consisting of 2 to 15 positions. In two cases there are 21 and 29 positions in a set. The myoglobins have strikingly high contribution of dispersed correlated positions. Out of 41 sets found,

(a)

| Position number and occurring residues | Correlation versus position 29 | |
|---|---|---|
| **29 [DEFIL]** | **L (37)** | **E (12)** |
| 6 [EGHKPSTW] | EGKST | W |
| 13 [-ADFIKLMPRSTV] | -AFILPRT | M |
| 40 [AEGILMV] | AILMV | E |
| 48 [KLMPQR] | KLMQ | R |

(b)

| Position number and occurring residues | Correlation versus position 13 | | |
|---|---|---|---|
| **13 [-ADFIKLMPRSTV]** | **L (11)** | **M (10)** | **A (8)** |
| 4 [-RSTVY] | V | -S | S |
| 5 [-KPST] | K | -S | S |
| 7 [AEGKP] | A | P | P |
| 11 [EFHIKLQRST] | T | EHQ | S |
| 21 [EFIKMQT] | T | Q | EQ |

(c)

| Position number and occurring residues | Correlation versus position 25 | |
|---|---|---|
| **25 [AEHMNQRSTV]** | **A (15)** | **V (9)** |
| 11 [EFHIKLQRST] | EFKLRS | HQ |
| 23 [HQRSV] | Q | R |
| 50 [FHIQRSVY] | HRS | FI |

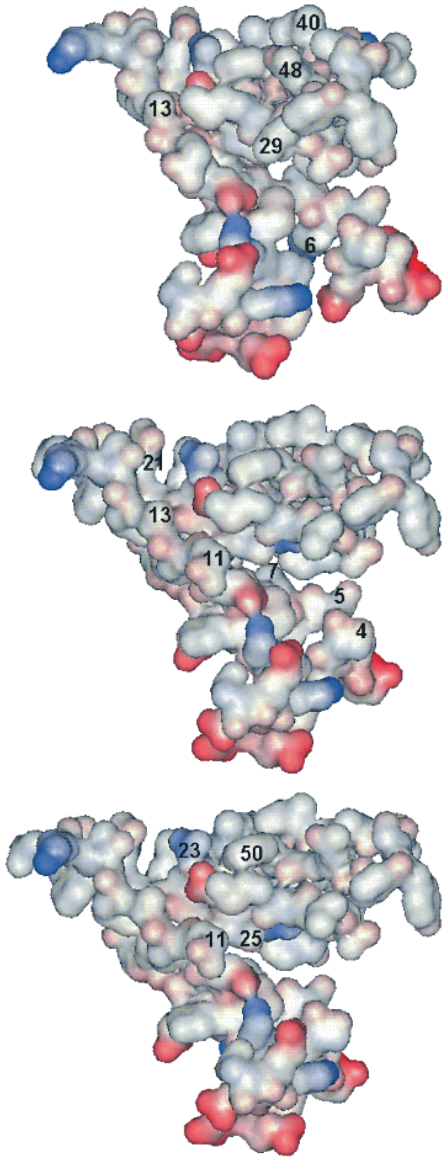



**Figure 4.** The three types of distribution of correlated positions present in Bowman-Birk inhibitor family: (a) the dispersed correlation, (b) the narrow correlation cluster, (c) the spot correlation cluster; the residue location and relative distribution is shown on tertiary structure of Bowman-Birk inhibitor from soybean (P01055)

23 are classified as dispersed, 9 can be assumed as narrow path clusters and 9 as wide-shaped clusters (Figure 8). According to the current outlook [6] one mutation in myoglobins is compensated by another in order to rescue the structural motif of the molecule. This hypothesis strongly suggests the correlation between the mutated residues that are spatial neighbors. The correlation is to preserve the seven-helical arrangement typical for myoglobins and proper antiparallel contacts between some of them. The results presented here are not fully concordant with that conclusion.

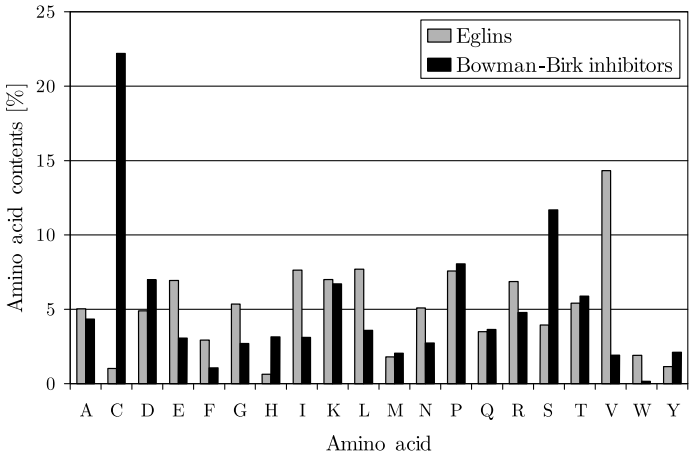



**Figure 5.** The relative amino acid composition of eglin-like proteins and Bowman-Birk inhibitors

More than half of the correlations in myoglobins are classified as dispersed, referring to distant residues that are not in direct contact and cannot interact spatially. Only about 44% of observed correlations may confirm the current hypothesis.

The results of lysozyme analysis (56 sequences) resemble those obtained for myoglobins. They show high similarity degree within the family. Almost all aligned positions yield consensus information (Figure 9). The positions revealing mutational correlation were grouped into 41 separate sets. Only nine of them contain positions of the catalytic site (or positions close to it with respect to the sequence distance). In 25 cases the correlation regards the positions distant from each other, in 16 sets the residues are clustered in limited area (Figure 10). Nine clusters form narrow paths. The mutational correlation among lysozyme sequences refers to 2–15 residues.

Although the general sequence outline of myoglobins and lysozymes is conservative (consensus sequence is well defined) the variability pattern also shows high diversity (Figure 11). There are the positions of high tolerance regarding the residue number (up to 10–11) in both families.

The correlations occurring among lysozymes confirm the general hypothesis of their significant role in intermolecular interaction only to some extent. Only a little more than 20% of observed correlations include the residues of the reactive site or their close neighbors. Most of the compensated mutations comprise the positions located far from the active center. Also, in more than 50% of cases the correlated residues are located very far from each other. It is very unlikely that these sets have any contribution to inner structure stabilization or intermolecular interaction.

It is still not known what are the evolutionary mechanisms of dispersed correlation in all analyzed protein families. Although it is known that residues not being in direct contact may strongly interact in indirect way, the frequent (almost 50%) occurrence of scattered residues in correlation sets, very far distance between these amino acids (12–25 Å) and lack of any other changes except for these residues does not confirm the model of simple indirect interaction effect. We suggest that their frequent occurrence in each family may be due to another differentiation mechanism from the hypotheses described so far.

| EGLIN-LIKE PROTEINS | | | |
|---|---|---|---|
| 8 | KR | 43 | -EILV |
| 9 | -ELNQRST | 44 | -DL |
| 10 | EFMQRST | 45 | -ANS |
| 11 | FW | 46 | DGM |
| 12 | P | 47 | GQSTV |
| 13 | EHQ | 48 | AFHINPV |
| 14 | LV | 49 | -W |
| 15 | CILV | 50 | -AFIV |
| 16 | EG | 51 | -T |
| 17 | ACKLMSTV | 52 | AEKLMQT |
| 18 | DGPRST | 53 | DEN |
| 19 | AGITV | 54 | EFILY |
| 20 | ADEKLS | 55 | DKLNR |
| 21 | ADEFKLQVY | 56 | CFILPY |
| 22 | A | 57 | DEKNQ |
| 23 | AEKMRV | 58 | R |
| 24 | AEGKQT | 59 | IV |
| 25 | -U1IKTVY | 60 | FR |
| 26 | FIV | 61 | ILV |
| 27 | EKLQT | 62 | FLWY |
| 28 | AEKLQRT | 63 | DNVY |
| 29 | DEHQ | 64 | ADHNT |
| 30 | KMNRY | 65 | -DEIKLPRV |
| 31 | PSV | 66 | -AGLNRS |
| 32 | DEKLNQRS | 67 | -DGNT |
| 33 | AILVY | 68 | -DFIKLNSTY |
| 34 | DEKQRST | 69 | IV |
| 35 | -AINV | 70 | ANTV |
| 36 | -E | 71 | DKNQRSZ |
| 37 | -V | 72 | AHIMPTV |
| 38 | -EHIQY | 73 | -ASV |
| 39 | -FILTV | 74 | -P |
| 40 | -LMSV | 75 | -AHKQRSTV |
| 41 | -P | 76 | IV |
| 42 | -EHIQRS | 77 | AGT |

| BOWMAN-BIRK INHIBITORS | | | |
|---|---|---|---|
| 3 | -DEKQST | 35 | ADET |
| 4 | -RSTVY | 36 | C |
| 5 | -KPST | 37 | DEKLNS |
| 6 | EGHKPSTW | 38 | ADEFGHKLRST |
| 7 | AEGKP | 39 | C |
| 8 | C | 40 | AEGILMV |
| 9 | C | 41 | CEKP |
| 10 | DNRS | 42 | ANRSTV |
| 11 | EFHIKLQRST | 43 | EFHKLRTVY |
| 12 | ACQ | 44 | DGS |
| 13 | -ADFIKLMPRSTV | 45 | DEFIMNQSY |
| 14 | C | 46 | -DPS |
| 15 | C | 47 | -AGPS |
| 16 | AKR | 48 | KLMPQR |
| 17 | S | 49 | CHR |
| 18 | DEFIKMNQR | 50 | FHIQRSVY |
| 19 | P | 51 | -I |
| 20 | AP | 52 | C |
| 21 | EFIKMQT | 53 | ABEFGLQTVY |
| 22 | C | 54 | DN |
| 23 | HQRSTV | 55 | -IMQTV |
| 24 | C | 56 | DHKNQTY |
| 25 | AEHMNQRSTV | 57 | -DHIKNRTV |
| 26 | DNQ | 58 | -FGY |
| 27 | -IKLMQTV | 59 | -CDI |
| 28 | GLRV | 60 | -HPTY |
| 29 | DEFIL | 61 | -ADEGKP |
| 30 | DEKNQRT | 62 | -AKPQS |
| 31 | -S | 63 | -MT |
| 32 | C | 64 | -C |
| 33 | AHPS | 65 | -DEHKNR |
| 34 | ADS | 66 | -DENPS |

**Figure 6.** The amino acids occurring at variable positions of eglin and Bowman-Birk family

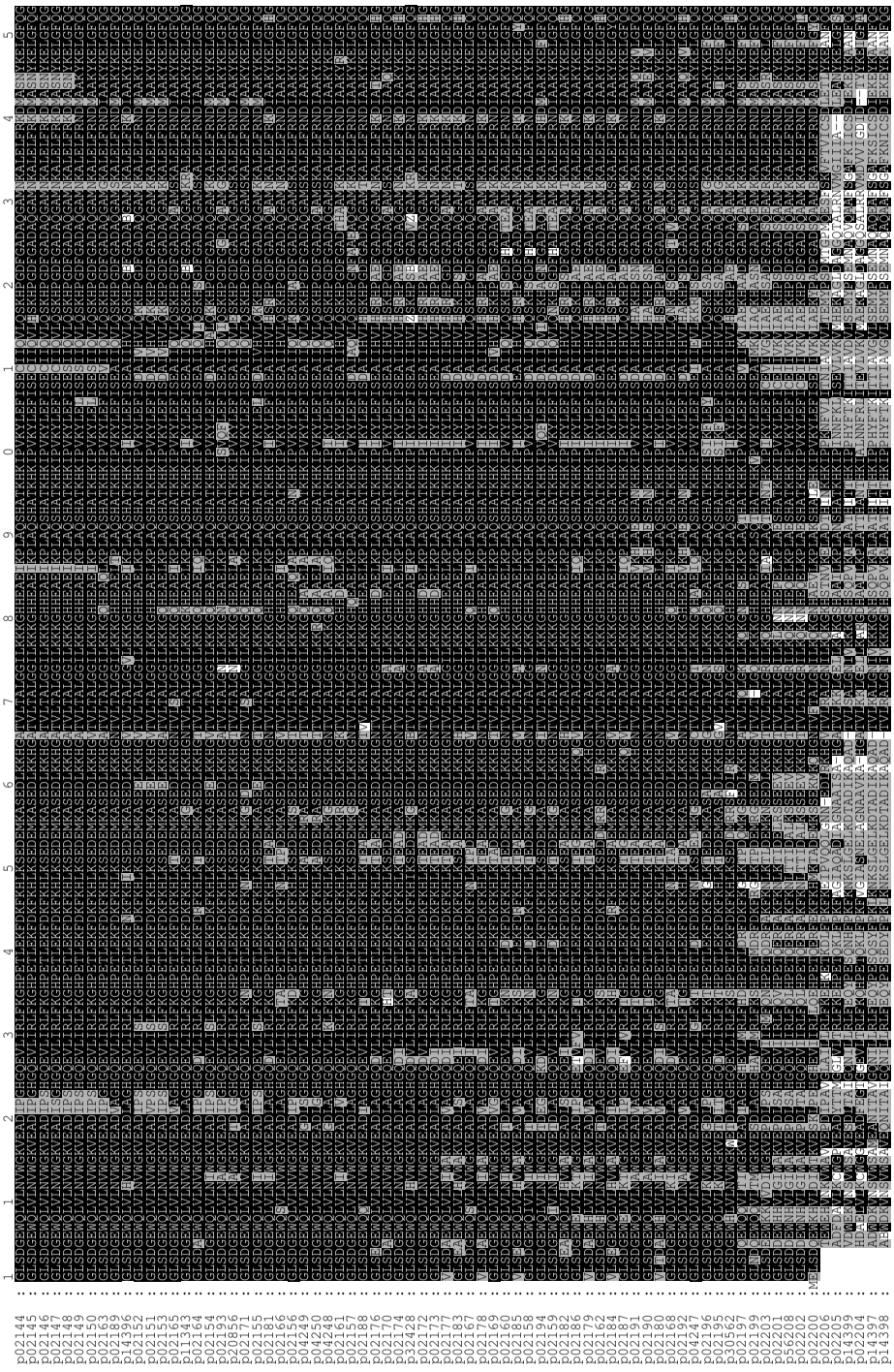
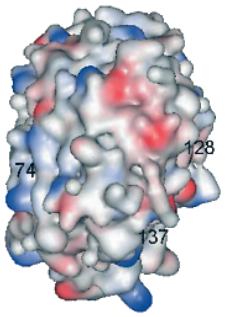


**Figure 7.** The multiple alignment of myoglobins, constructed according to the genetic semihomology algorithm; the conservative consensus residues are shown as white letters on black background; the gray background indicates genetic semihomology between aligned residues

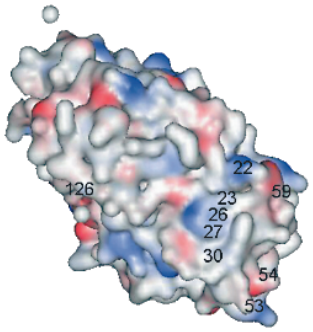
(a)

| Position number and occurring residues | Correlation versus position 74 | | |
|---|---|---|---|
| **74 [AGPQST]** | **A (6)** | **G (49)** | **N (9)** |
| 128 [ABEHQ] | Q | BEHQ | Q |
| 137 [ILNSV] | L | L | ILNSV |




(b)

| Position number and occurring residues | Correlation versus position 23 | | |
|---|---|---|---|
| **23 [AEGST]** | **A (7)** | **G (55)** | **S (10)** |
| 22 [AEGPSTV] | PST | AEGPSTV | P |
| 26 [EGHKLQ] | LQ | EGHQK | Q |
| 27 [ADEFLNT] | AEN | ADEFT | E |
| 30 [ILMTV] | IL | IMTV | I |
| 53 [ADEGQ] | DEQ | ADEG | D |
| 54 [ADEILQ] | AEL | DELQ | E |
| 59 [ADEF] | DE | ADEF | E |
| 128 [ABEHQ] | Q | ABEHQ | Q |




(c)

| Position number and occurring residues | Correlation versus position 127 | |
|---|---|---|
| **127 [AMSTV]** | **A (58)** | **S (7)** |
| 27 [ADEFLNT] | ADEFNT | E |
| 31 [GKRS] | GKRS | R |
| 78 [AKLQ] | K | ALQ |
| 109 [DEGNT] | DEGT | E |
| 116 [AEHKQST] | AEHKQS | A |
| 117 [AEKNQS] | AEKQS | E |
| 122 [BDEN] | BDEN | D |

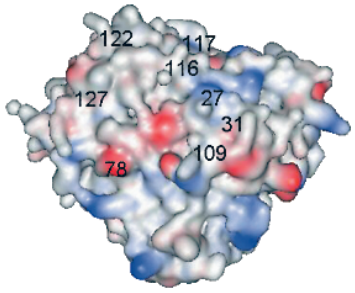



**Figure 8.** The three types of distribution of correlated positions present in myoglobins: (a) the dispersed correlation, (b) the narrow correlation cluster, (c) the spot correlation cluster; the sequence numbering in tables is as in Figure 7; the residue location and relative distribution is shown on tertiary structure of human myoglobin (P0244, pdb1bzp)

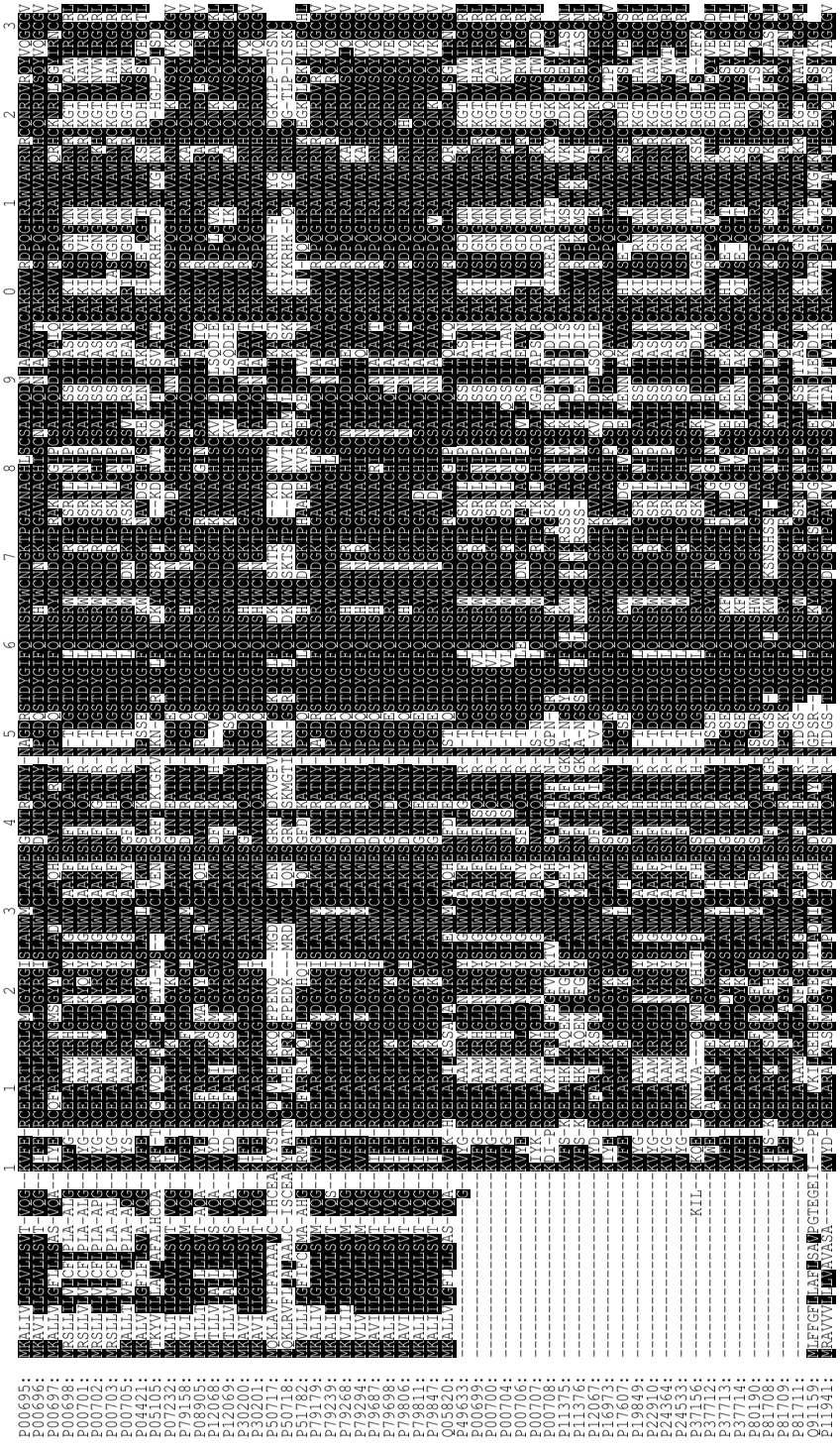
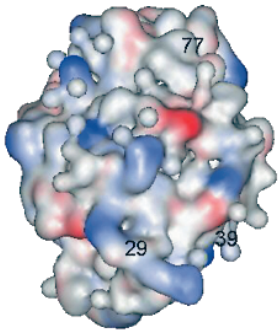


**Figure 9.** The multiple alignment of lysozymes, constructed according to the genetic semihomology algorithm; the conservative and frequently occurring consensus residues are shown as white letters on black background

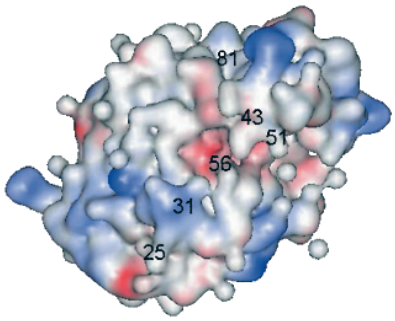
(a)

| Position number and occurring residues | Correlation versus position 80 | | |
|---|---|---|---|
| **80 [GHKNR]** | **G (7)** | **H (31)** | **N (16)** |
| 30 [ILMV] | MV | ILMV | V |
| 40 [DFKNR] | DN | N | FKNR |




(b)

| Position number and occurring residues | Correlation versus position 59 | |
|---|---|---|
| **59 [FL]** | **F (38)** | **L (18)** |
| 26 [-ILMV] | -ILMV | L |
| 33 [AISTV] | AISTV | A |
| 44 [FIMRTVY] | FIMRTVY | T |
| 54 [-KRSTY] | -KRSTY | T |
| 84 [AKNRS] | AKNRS | S |




(c)

| Position number and occurring residues | Correlation versus position 126 | |
|---|---|---|
| **126 [-DWY]** | **W (16)** | **Y (36)** |
| 13 [-ILM] | M | IL |
| 15 [-AEKNQRS] | R | AEKNQRS |
| 20 [DEGKN] | KN | DG |
| 27 [-AEGPR] | G | AEP |
| 44 [FIMRTVY] | T | FIRTY |
| 46 [GHKNPRTY] | R | GHNRY |
| 105 [-AGHPQRV] | GRV | -APQ |
| 109 [DGKNQRST] | N | GKRST |
| 121 [-HKQRT] | T | HKQR |

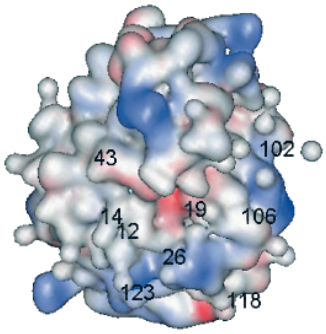



**Figure 10.** The three types of distribution of correlated positions present in lysozymes: (a) the dispersed correlation, (b) the narrow correlation cluster, (c) the spot correlation cluster; the sequence in tables numbering is as in Figure 9; the residue location and relative distribution is shown on tertiary structure of lysozyme from rat (P00697, pdb5lyz)

## *Acknowledgements*

The calculations were partially done on the computers in the Wroclaw Center of Networking and Supercomputing, grant no. 14/97. This work was also supported from BST2001 grant no. 115/E-343/S/BST701/2001/ICM.

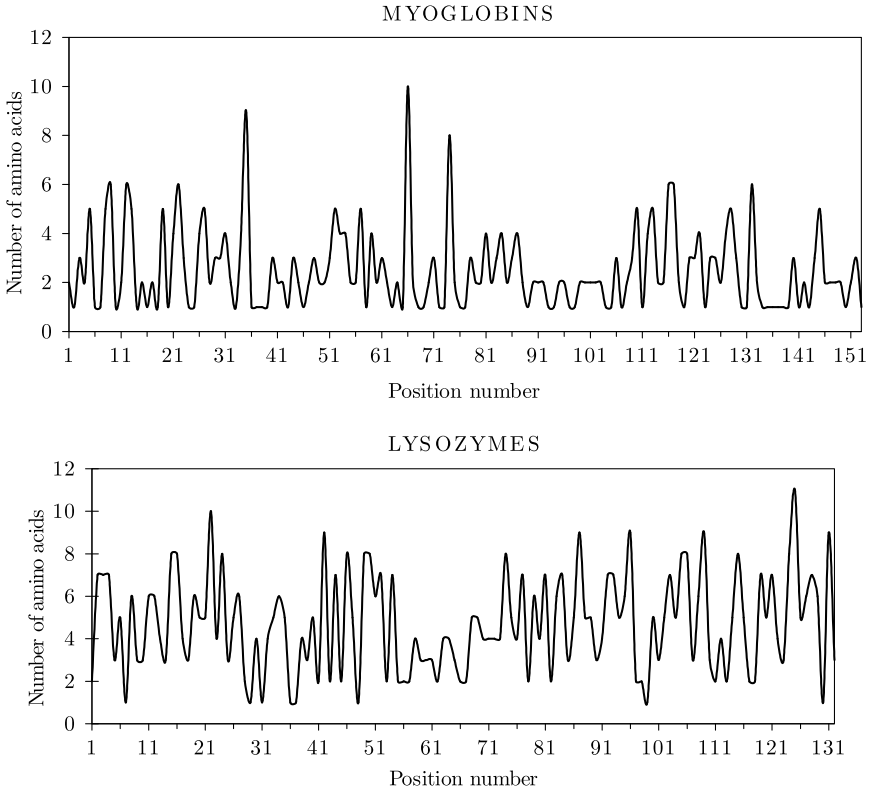



**Figure 11.** The position variability patterns of myoglobins and lysozymes

## *References*

[1] Kimura M 1983 *The Neutral Theory of Molecular Evolution*, Cambridge University Press, Cambridge
[2] Laskowski Jr M, Kato I, Kohr W J, Park S J, Tashiro H E and Whatley H E 1988 *Cold Spring Harbor Symp. Quant. Biol.* **52** 545
[3] Laskowski Jr M and Fitch W M 1989 *The Hierarchy of Life* (Fernholm B, Bremer K and Jornvall H, Eds.), Elsevier Science Publishers B. V., Chapter 27, pp. 371–387
[4] Hill R E and Hastie N D 1987 *Nature* **326** 96
[5] Pollock D D and Taylor W R 1997 *Protein Engng.* **10** 647
[6] Neher E 1994 *Proc. Natl. Acad. Sci. USA* **91** 98
[7] Pazos F, Helmer-Citterich M, Ausiello G and Valencia A 1997 *J. Mol. Biol.* **271** 511
[8] Apweiler R, Gateau A, Contrino S, Martin M J, Junker V, O'Donovan C, Lang F, Mitaritonna N, Kappus S and Bairoch A 1997 *Proc. 5$^{th}$ Int. Conf. on Intelligent Systems for Molecular Biology*, AAAI Press, Menlo Park, pp. 33–43
[9] Bairoch A and Apweiler R 1997 *J. Mol. Med.* **75** 312
[10] Bairoch A and Apweiler R 1999 *Nucleic Acids Res.* **27** 49
[11] Bairoch A 1997 *Proteome Research: New Frontiers in Functional Genomics* (Wilkins M R, Williams K L, Appel R D and Hochstrasser D H, Eds.), Springer-Verlag, Heidelberg, pp. 93–132
[12] Bairoch A and Apweiler R 2000 *Nucleic Acids Res.* **28** 45
[13] Altschul S F, Gish W, Miller W, Myers E W and Lipman D J 1990 *J. Mol. Biol.* **215** 403
[14] Leluk J 1998 *Computers Chem.* **22** 123
[15] Leluk J 2000 *Biosystems* **56** 83
[16] Leluk J 2000 *Computers Chem.* **24** 659