Elsevier Editorial System(tm) for Carbohydrate Research
Manuscript Draft

Manuscript Number: CAR-D-10-00518R1

Title: Computer-aided subsite mapping of α-amylases

Article Type: Full Length Article

Section/Category: Biochemistry and Enzymes

Keywords: alpha-amylase; subsite mapping; binding energy; bond cleavage frequency; molecular modeling

Corresponding Author: Dr. Péter Bagossi, Ph.D.

Corresponding Author's Institution: University of Debrecen

First Author: János A Mótyán

Order of Authors: János A Mótyán; Gyöngyi Gyémánt, Ph.D.; János Harangi, Ph.D.; Péter Bagossi, Ph.D.

Abstract: Subsite mapping is a crucial procedure in characterization of α-amylases (EC 3.2.1.1) which are extensively used in starch based industries and in diagnosis of pancreatic and salivary glands disorders. A computer-aided method has been developed for subsite mapping of α-amylases, which substitutes the difficult, expensive and time-consuming experimental determination of action patterns to crystal structures based energy calculations. Interaction energies between enzymes and carbohydrate substrates were calculated after short energy minimization by a molecular mechanics program. A training set of wild type and mutant amylases with known experimental action patterns of 13 enzymes of wide range of origin was used to set up the procedure. Calculations for training set resulted in good correlation in case of subsite binding energies (r^2 = 0.827–0.929) and bond cleavage frequencies (r^2 = 0.727–0.835). A set of eight novel barley amylase 1 mutants was used to test our model. Subsite binding energies were predicted with r^2 = 0.502 correlation coefficient, while bond cleavage frequency prediction resulted in r^2 = 0.538. Our computer-aided procedure may supplement the experimental subsite mapping methods to predict and understand characteristic features of α-amylases.

Reviewer #1: Recommendation: Minor Revision

This manuscript reports the use of computation to determine bond cleavage frequencies and subsite binding energies from crystal structures of a number of α-amylases. This is an interesting choice of enzymes, as α-amylases are extensively used industrially and a great number of members of their family (GH13) have been sequenced and subjected to crystal structure determination. They are also endo-hydrolases, so different positions of starch chains can be cleaved. The argument is made that use of computation saves a great deal of time and money compared to obtaining the same data experimentally. This is undoubtedly true; certainly subsite mapping by experimental means is a long and hard process, as we can attest. On the other hand, computation must always take a back seat to experimentation if the latter can be used to obtain the data. After all, computation is subject to errors introduced by incomplete and inaccurate modeling of already existing experimental data. Comparison of experimental and computational data in this manuscript confirms the statement just above – the agreement between them is often rather poor.

We completely agree with the reviewer therefore we emphasized throughout the manuscript that our method may complement experimental data but not substitute them. One of the reasons to choose relatively limited data sets was that we wanted to develop and apply our method to close relatives of enzymes where accuracy of modeling was usually higher than those of distant relatives. If the first try produce satisfactory results, we may extend the scope of the method later. Third, we also predicted the possible reasons where our method was failed and a special attention was needed during the work.

Despite this, the manuscript is probably worth publishing as a first attempt to complement experimental subsite mapping. The English here is in general clear but is certainly the product of non-native English speakers. It will need some rewriting and correcting by a production editor.

Following are specific comments as I read the manuscript:

p. 3, line 11: Are plants not higher organisms?

The sentence was corrected to the following: "They can be found in microorganisms, plants, animals and human where they play a dominant role in carbohydrate metabolism."

p. 3, line 44: Is Domain C a carbohydrate-binding module? If so, what CBM families are found in the α-amylases studied here?

Domain C is not a carbohydrate-binding module (CBM) and enzymes of our data sets do not contain any well-defined separate CBM or starch-binding domain (SBD). α-Amylases belonging to GH-13 family are multidomain proteins that contain several characteristic domains such as the obligatory catalytic domain A and most of them possess a domain B and domain C. Some enzymes in GH 13 contain one or two additional all-β domains: domain D and/or domain E, at the C-terminal end, following the domain C (Janecek et al., 2003, Eur. J. Biochem. 270, 635–645). The function of domain D is unknown; however, domain E is referred to as SBD (the term of SBD still often used in the amylase research instead of the more "official" CBM) which is a distinct sequence-structural module that improves the efficiency of an amylolytic enzyme on raw starch. Four- and five-domain members of GH-13 can be referred to generally as the SBD-containing hydrolases and they can be classified into various CBM families.

In contrast to the fact that none of the enzymes studied here contain domain D and domain E, they may contain secondary binding sites for carbohydrates in domains A, B or C. For example, the barley α-amylase 1 (AMY1) provides two binding sites in addition to the catalytic cleft: a starch granule binding site within the catalytic domain A and a so-called sugar tongs within domain C (Nielsen et al., 2008, Biocatal. Biotransf., 26, 59-67). HSA has also surface binding sites located far from the active site (Ramasubbu et al., 2003, J. Mol. Biol., 325, 1061-1076). These enzymes are capable of binding and digesting raw starch without a specialized functional domain (e.g. SBD) in their sequence and structure (Janecek et al., 2003, Eur. J. Biochem. 270, 635–645).

The manuscript was modified accordingly: "Besides domain A, B and C, the type and the number of extra domains such as domain D and/or domain E located at the C-terminus show wide variety within the α-amylase family. The function of domain D is unknown; however, domain E is referred to as carbohydrate-binding module (CBM) or starch-binding domain (SBD) which is a distinct sequence-structural module that improves the efficiency of an amylolytic enzyme on raw starch. It should be noted that none of the enzymes studied here contain domain D and/or domain E."

p. 3, lines 53–56: Is it not true that α-amylases mainly act on α-(1,4)-linked glucans?

Yes, it is true. Families 13, 70 and 77 of glycoside hydrolases contain structurally and functionally related enzymes catalyzing hydrolysis or transglycosylation of α-linked glucans, with retention of anomeric configuration. Alpha-amylases catalyze the endohydrolysis of (1→4)-α-D-glycosidic linkages in polysaccharides containing three or more (1→4)-α-linked

D-glucose units and they act on starch, glycogen and related polysaccharides and oligosaccharides in a random manner. All these information can be found in the first two paragraphs of the manuscript. Nevertheless we rephrased the above mention sentence of the manuscript to the following: "Members of the α-amylase family act on several substrates (starch, glycogen, oligosaccharides) and the only shared structural feature of substrates is an α-(1-4)-linked glucose residue that should bind in the first aglycone subsite near the scissile bond."

p. 5, line 35: It would be helpful to note that these PDB structures were of maltose and maltoheptaose crystallized with α-amylases.

The sentence was modified to contain this information: "Structures of the maltooligosaccharide substrates were modeled based on the crystal structures of maltose (PDB code: 2GVY), maltoheptaose (PDB code: 1RP8) and a substrate-analogue inhibitor acarbose (PDB code: 1MFU, 1RPK, 1E3Z, 1OSE) complexed with various alpha-amylases."

p. 5, line 42: What does "bumped" mean here?

Crystal structures used for modeling of enzymes of our data sets contained shorter substrate or inhibitor than that of ours, therefore water molecules which occupied an "empty space" in the original x-ray structure may have a very close contact to the long substrate of the model (distance is much shorter than the sum of the van der Waals radius of the two atoms). Not surprisingly, it happened that oxygen atom of a hydroxyl group of our model substrate occupied approximately the same position as the oxygen atom of a water molecule in the crystal structure. In this case we removed that water molecule to avoid distortion caused by the huge initial energy during the minimization.

p. 7, line 23: Except for the list of abbreviations, this is the first time that $E_{transf.}$ has been mentioned. How is it calculated?

The sentence was modified to explain the calculation: "After regression analysis, E_{Sybyl} data were linearly transformed to the scale of E_{SUMA} data with the values of intercept and slope of the best fit line to get $E_{Transf.}$ (Fig. 2A and 2B) for each enzyme group."

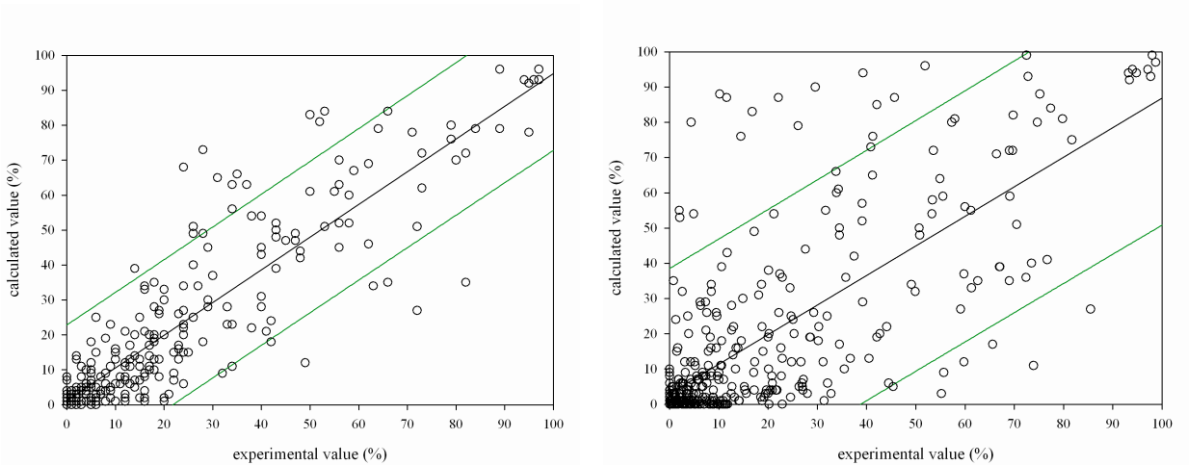
p. 8, line 31: A table or figure showing calculated BCF's would be helpful.

The table containing all BCF data would be too large here because BCFs of series of oligomeric substrates used to calculate a single set of binding energies. Furthermore, the BCF data will be published soon in a separate paper (Mori et al., in preparation). We may show an

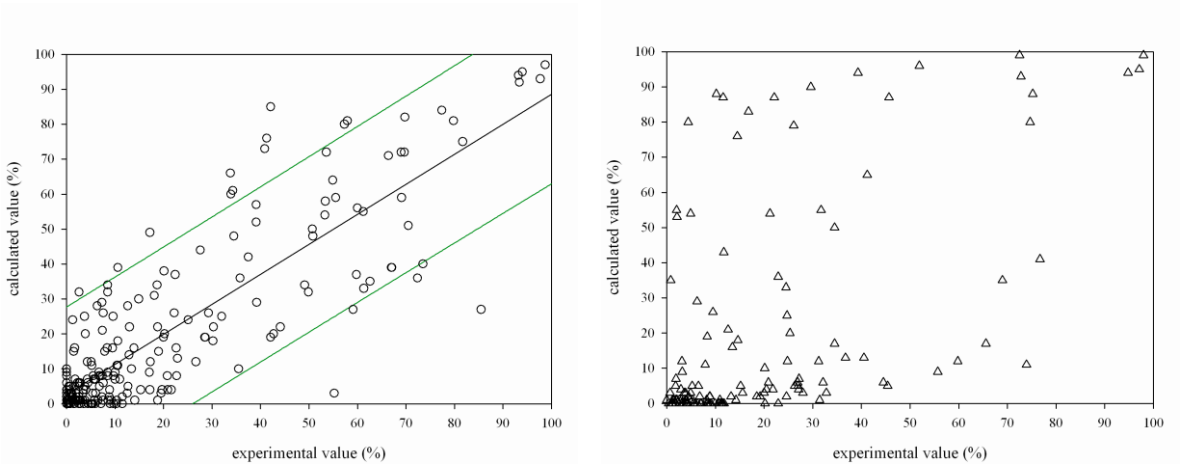
example for the correlation of the experimental and the calculated data, but these graph are still contained large number of points and they look a little bit messy. On the other hand, these graphs show basically the same qualitative message as the Fig. 2 but if the Reviewer suggests including any of them into the manuscript, we will do it.

New figure? Correlation of experimental and calculated BCF values of wild-type and mutant AMY1 enzymes: a) training set (r^2 = 0.801) and b) test set (r^2 = 0.538). Majority of points located outside of the 95 % prediction interval range are belonging to mutant containing charged residue. To show this, the test set is divided into two classes: c) data of test set without those of mutant containing charged residues (r^2 = 0.728) and d) data of charged mutant only.

a) b)



c) d)



References: Journal titles should be abbreviated when they are of more than one word.

References are reformatted.

Table 1: What is being correlated here?

Subsite binding energies were recalculated using our subsite models and this table shows the excellent correlation between the published energy values and our recalculated ones. This was explained only in the text therefore legend of Table 1 was modified to the following: "Original and modified subsite models of the studied enzymes and the correlation between the subsite binding energies of the original publications and the recalculated values based on subsite models of this work."

Table 2: Individual subsite binding energies are very different between enzymes. Can a statement be made in the text linking this observation to the chain lengths of characteristic products of these enzymes?

The following sentences were inserted into the Introduction: "The distribution of the high and low affinity binding subsites and the barrier subsites within the active site determines the hydrolytic efficiency on series of substrates with various length: high-affinity subsites next to the cleavage sites (−2 through +2) allow effective cleavage of shorter substrates, while distal high-affinity subsites control the action on longer substrates. Chain length and concentration of characteristic products can be determined from the BCF tables and energy contribution of each subsite to the total binding energy can be calculated."

Table 3: The table legend mentions energies but not BCF's.

Legend of Table 3 was corrected to the following: "Parameters of the best fitted line of linear regression analysis of E_{SUMA} and E_{Sybyl} values (middle panel) and the experimental and calculated values of BCFs (right panel) (m − slope, b − intercept, r^2 − square of correlation coefficient)."

Table 4: This table is not mentioned until the Conclusions. It should have been mentioned earlier.

The tables were renumbered according to the order of appearance (earlier Table 4 now is Table 1) and the following text was inserted into the first paragraph of the Result and discussion: "Wild-type enzymes were chosen based on the existence of experimental BCF

values as well as available crystal structure. In contrast to the various degree of sequence identity (11-86 %) between the wild-type enzymes (Table 1), they were all classified into the glycoside hydrolase family 13 and showed high degree of structural conservation."

Figure 2A: A plot of kcal/mol vs. kJ/mol does not allow the reader to easily determine whether the slope of the regression line is unity.

Fig. 2A shows only the raw data and it is not aimed to demonstrate that the slope is unity. Even if the kcal/mol is converted to kJ/mol, nobody can except that the absolute scale of the experimental and the calculated energy values would be the same. During the process of the calculation several assumptions were made, several factors were neglected, only partial minimizations were done, etc. This was the reason that the transformation step was needed and the unit conversion was also included in this step.

Reviewer #2: Recommendation: Minor Revision

The manuscript by Motyan et al. reports on an application of their SUMA algorithm to the problem of subsite mapping in alpha-amylases. The paper is relatively well written and addresses a moderately interesting problem in the field. The authors have collected a nice training and test set of data, and then proceed to apply their methodology to optimize empirical parameters to fit the data. This is a very common practice and methodologically, this paper does not present anything new. As such, the most interesting result in the paper is the model specifically trained for alpha-amylases. There are major weaknesses in the paper that should however be addressed before publishing.

1.) The size of the training and test sets are still small. Ideally, the authors should try to at least double the size of the training set to be able to perform more comprehensive cross-validation. The fits are in fact not cross-validated, and the authors do not at all address whether they may in fact have overfit their model.

This was our first attempt to study on in silico subsite maps of carbohydrate-modifying enzymes and similar work had not been published in the literature, therefore we decided to limit our study to the relatively well characterized, structurally similar, catalytically uniform group of alpha-amylases. They still represented a wide range of enzymes

of various species from bacteria to human: the sequence identity values were 11-86 % (Table 1 of the revised manuscript).

We must limit our training set to the enzymes which experimentally determined BCF and crystal structure were both existed. We included all published experimental data in our training set except two: BCF and subsite map of rice amylase were published (Gyémánt et al., Eur. J. Biochem., 2002, 269, 5157-5162) but its crystal structure has not been determined yet and crystal structure of maltogenic α-amylase of *Bacillus stearothermophilus* was determined but BCF data has not been published yet.

We admit that the size of our data set did not allow to handle all aspects in statistically satisfactory way but we think that our results and conclusions are correct and not go beyond the scope of this study.

2.) Why are the authors focusing only on alpha-amylases? Why not perform this analysis on all amylases, which would greatly increase the size of their training set and allow for much more 10-fold cross validation studies.

The most known amylolytic enzymes are α-amylase (EC 3.2.1.1), β-amylase (EC 3.2.1.2) and glucoamylase (EC 3.2.1.3); however, they differ from each other in their primary and tertiary structures, their catalytic machineries and reaction mechanisms. They have therefore classified into different families: GH13 - α-amylases, GH14 - β-amylases, GH15 - glucoamylases. It may possible that our method can be applied not only to GH13 but further families too; however, it needs further studies and it is beyond the scope of this manuscript.

SUMA program was used in several step of our method and this program was developed and validated only for α-amylases, as its full name in the Abbreviations and in the text indicates: SUbsite Mapping of α-Amylases.

Please also see the answer for the first question of this reviewer.

3.) The authors are using the AMBER force field, yet there are far more accurate force-fields for sugars, e.g. glycam. The authors should evaluate how using different force-fields affect their results. Furthermore, the authors should also evaluate whether using explicit solvent in their simulations improves (or perhaps worsens) their fits to the experimental data.

Several force fields available in Sybyl program: Kollman_All_Atom, Amber95 and Amber7_FF99 were tested, together with an in-house implementation of Glycam06 force field. Amber7_FF99 was chosen because its performance was the best in our hand. Furthermore, wide-variety of values of dielectric constant (1, 2, 3, 4 and 8), non-bonded

cutoff (8, 10 and 12), number of iterations (20-100 by steps of 20 and 100-1000 by steps of 100) and the width of minimized shell (1-12 angstrom) of atoms around the substrate and the mutated residues in the initial energy minimization were also probed to find a good combination of parameters and it was turned out that these factors also significantly modified the results. Several, but not all possible combinations were tested because it would require large amount of computational power, therefore extensive discussion on the parameter choice was neglected. One of our aims was to build a fast procedure; therefore explicit molecules for bulk solvent and extensive minimization were avoided. We think that the readers of Carbohydrate Research are more interesting in the chemical/biochemical aspects of our method rather than the computational details of the parameter choice. The text of the manuscript was modified accordingly.

4.) *The experimental data used to fit have been collected under a wide-variety of solution conditions. The authors need to discuss and assess how experimental error may in fact be limiting the resolution of their methodology.*

The wild-type enzymes were measured at various reaction conditions; however, they were measured at the optimal condition of each enzyme and this situation may be more appropriate for determination of bond cleavage frequencies and subsite binding energies than the choice of one reaction condition which is optimal for only one enzyme and suboptimal for the all others. Definitive conclusion of the effect of each reaction condition (e.g. pH, temperature or buffer composition) can be drawn only after extensive experimentation. Mutant enzymes were measured at the same reaction condition than their wild-type enzymes based on the assumption that their features were changed moderately. The same assumption was also necessary for the theoretical calculations. Furthermore, the effect of different reaction conditions could be eliminated in a large extent with our energy transformation procedure which was group-specific. Enzyme groups were based on each wild-type enzymes and mutants belonged to the corresponding wild-type group.

Differences of the calculated subsite binding energies were below 0.4 kJ/mol in case of two independent experimental measurements of BCF, as we wrote in the Materials and methods. In a recent publication of us it was found that the same type of error was below 0.6 kJ/mol (Nielsen et al., 2009, Biochemistry 48, 7686–7697).

5a.) *It is not clear how their results will be shared with the world. Can we access their parameters on a web server to test with other amylases?*

We have no license to allow online calculation with Sybyl program; therefore a dedicated web server cannot be setup for this kind of calculation. However, in this manuscript we publish all necessary parameters to repeat our calculations or setup a similar procedure with different molecular mechanics core, data sets, subsite models or target enzymes.

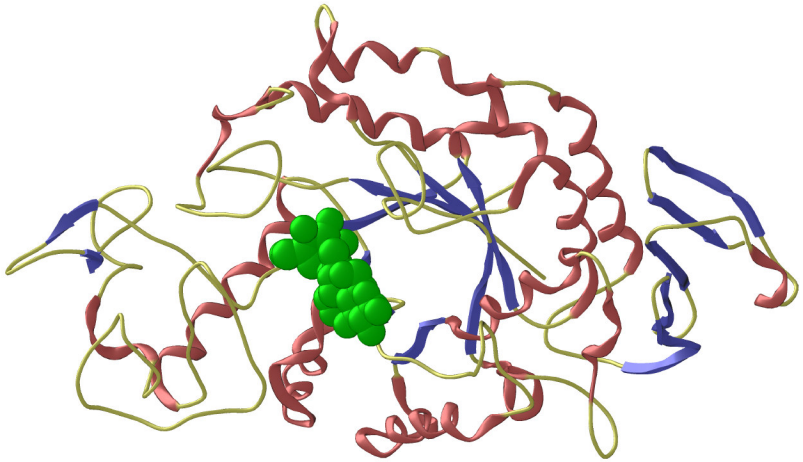
5b.) *Also, the authors should make their compiled literature search data available as a spreadsheet and cite the papers they obtained the experimental data from.*

This question had already been addressed in Table 1 and Table 2 of the original manuscript. Last column of the Table 1 contains citations for all publications had been used for creation our data sets. We do not want to republish original data of those papers, however, all recalculated data based on our subsite models can be found in Table 2 (rows of E_{SUMA}) together with values calculated by our novel procedure (rows of $E_{Transf.}$).

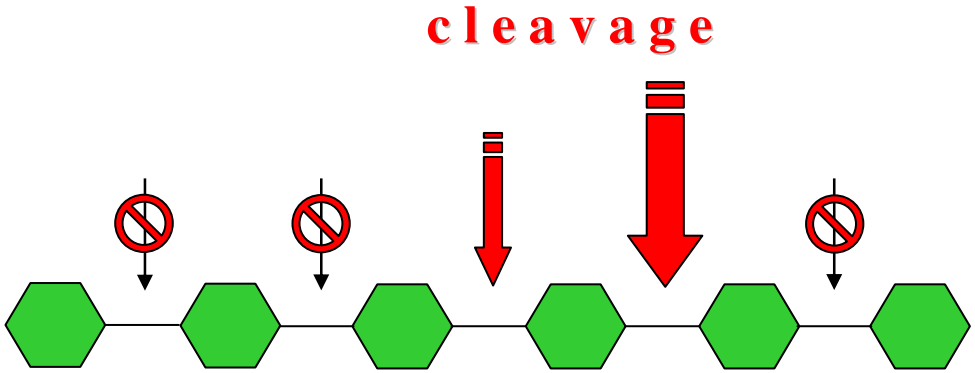
*Graphical Abstract

Click here to download Graphical Abstract: GraphicalAbstract-Bagossi.pdf

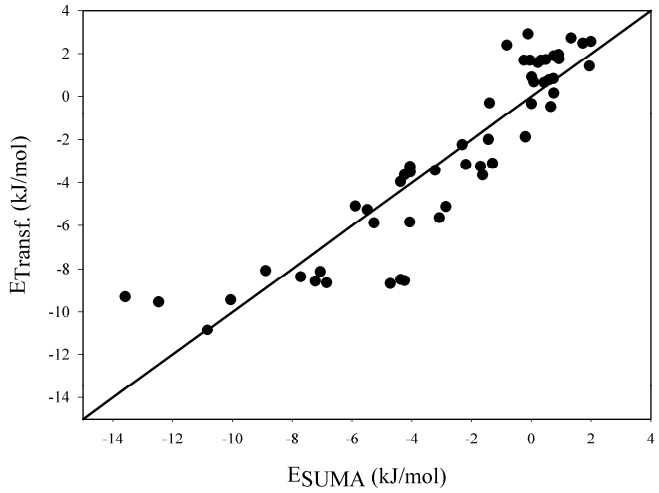
α-amylase



Energy calculation on xray structures or homologous models



Prediction of unknown subsite binding energies and bond cleavage frequencies



Training and test on experimental subsite maps

Computer-aided subsite mapping of α-amylases

János A. Mótyán[a,b], Gyöngyi Gyémánt[b,c], János Harangi[b,d], Péter Bagossi[a,*]

[a] Department of Biochemistry and Molecular Biology, Faculty of Medicine, Medical and Health Science Center, University of Debrecen, Debrecen, Hungary H-4010

[b] Department of Biochemistry, Faculty of Sciences and Technology, University of Debrecen, Debrecen, Hungary H-4010

[c] Present address: Institute of Inorganic and Analytical Chemistry, Faculty of Sciences and Technology, University of Debrecen, Debrecen, Hungary H-4010

[d] Present address: Department of Computer Science and Systems Technology, Faculty of Information Technology, University of Pannonia, Veszprém, Hungary H-8200

*** Corresponding author:** Péter Bagossi, Department of Biochemistry and Molecular Biology, Medical and Health Science Center, Faculty of Medicine, University of Debrecen, H-4012 Debrecen, P.O.Box 6, HUNGARY, Phone: (+36-52) 416-432, Fax: (+36-52) 314-989, Email: bagossi@med.unideb.hu

Email addresses: motyan@puma.unideb.hu (JM), gyemant@puma.unideb.hu (GG), harangi@dcs.uni-pannon.hu (JH), bagossi@med.unideb.hu (PB)

Abstract

Subsite mapping is a crucial procedure in characterization of α-amylases (EC 3.2.1.1) which are extensively used in starch based industries and in diagnosis of pancreatic and salivary glands disorders. A computer-aided method has been developed for subsite mapping of α-amylases, which substitutes the difficult, expensive and time-consuming experimental determination of action patterns to crystal structures based energy calculations. Interaction energies between enzymes and carbohydrate substrates were calculated after short energy minimization by a molecular mechanics program. A training set of wild type and mutant amylases with known experimental action patterns of 13 enzymes of wide range of origin was used to set up the procedure. Calculations for training set resulted in good correlation in case of subsite binding energies (r^2 = 0.827−0.929) and bond cleavage frequencies (r^2 = 0.727−0.835). A set of eight novel barley amylase 1 mutants was used to test our model. Subsite binding energies were predicted with r^2 = 0.502 correlation coefficient, while bond cleavage frequency prediction resulted in r^2 = 0.538. Our computer-aided procedure may supplement the experimental subsite mapping methods to predict and understand characteristic features of α-amylases.

Keywords:

alpha-amylase; subsite mapping; binding energy; bond cleavage frequency; molecular modeling

Abbreviations:

AMY1, barley α-amylase 1 isoenzyme; AMY2, barley α-amylase 2 isoenzyme; BCF, bond cleavage frequency; BAA, *Bacillus amyloliquefaciens* α-amylase; CNP, 2-chloro-4-nitrophenyl; DP, degree of polymerization; E_{SUMA}, subsite binding energy calculated from experimental BCF data by SUMA program; E_{Sybyl}, subsite binding energy calculated by Sybyl program; $E_{Transf.}$, subsite binding energy after linear transformation of E_{Sybyl}; G_n, maltooligosaccharide of *n* glucopyranoside units; HSA, human salivary α-amylase; MOS, maltooligosaccharide; PPA, porcine pancreatic α-amylase; SUMA, SUbsite Mapping of α-Amylases; TAA, Taka-amylase A (*Aspergillus oryzae* α-amylase).

1. Introduction

α-Amylases (α-1,4-glucan-4-glucanohydrolases; EC 3.2.1.1) constitute a family of endo-amylases catalyzing the cleavage of α-(1-4) glycosidic bonds in starch and related carbohydrates with retention of α-anomeric configuration in the products. They can be found in microorganisms, plants, animals and human where they play a dominant role in carbohydrate metabolism. Amylases are widely used enzymes in starch-based industries, play important role in processes of food and paper industry, textile desizing, detergent applications, sanitary waste treatment, starch liquefaction and saccharification and used to increase the digestibility of animal feed[1]. Both human salivary and pancreatic amylases are extensively studied enzymes from the viewpoint of clinical chemistry because of their importance in diagnosis of pancreatic and salivary glands disorders[2].

Carbohydrate-active enzymes are classified based on their sequence and structure similarities and grouped into more than hundred families according to the CAZy database[3]. α-Amylases belong to glycoside hydrolase family 13 (GH 13) and together with glycoside hydrolase families 70 and 77 constitute the clan H of glycoside hydrolases (GH-H)[3]. The three-dimensional structures of α-amylases mostly contain three domains (Fig. 1). The catalytic domain (domain A) adopts a central $(\beta/\alpha)_8$-barrel structure where the barrel of eight parallel β-strands are surrounded by eight helices. In most cases domain A occurs at the N-terminal end of the protein. Domain B is a small protruding loop located between the third β-strand and α-helix in the $(\beta/\alpha)_8$-barrel. The substrate binding site is located in the cleft at the interface of domain A and B and substrate binding subsites are made from the side chains of residues located on the loops at the C-terminal ends of β-strands of barrel structure[4]. The architecture of the domain B varies between amylases and it is the major determinant of differences of substrate specificities[5]. Most members of the family carry domain C, a C-terminal anti-parallel β-sheet composed of 5−10 strands following the catalytic $(\beta/\alpha)_8$-barrel and it plays an important role in stability/folding of the protein and in substrate binding[4].

Besides domain A, B and C, the type and the number of extra domains such as domain D and/or domain E located at the C-terminus show wide variety within the α-amylase family[6, 7]. The function of domain D is unknown, however, domain E is referred to as carbohydrate-binding module (CBM) or starch-binding domain (SBD) which is a distinct sequence-structural module that improves the efficiency of an amylolytic enzyme on raw starch. It should be noted that none of the enzymes studied here contain domain D and/or domain E.

Members of the α-amylase family act on several substrates (starch, glycogen, oligosaccharides) and the only shared structural feature of substrates is an α-(1-4)-linked glucose residue that should bind in the first aglycone subsite near the scissile bond[4]. Structure-based sequence alignment of key active site residues showed strong similarity between the members of GH13 family[8] but the β-α loops of domain A shows great variety between α-amylases, therefore the architecture of the substrate binding cleft (the number and nature of substrate binding subsites) is a specific feature of the particular enzymes[4]. The nomenclature for sugar-binding subsites of glycosyl hydrolases was introduced by Davies and co-workers[9]: subsites are labeled from the catalytic site, with negative numbers for subsites to the left (non-reducing end side, glycone binding sites) and positive numbers to the right (reducing end side, aglycone binding sites). Each subsite interacts with one glucose residue of the substrate. Nomenclature of subsites and schematic representation of substrate binding sites of enzymes used here are summarized on Fig. 1.

Subsite mapping is the process of quantifying the subsite model. Modified, low-molecular weight oligosaccharides are appropriate substrates for amylases to elucidate the number of subsites in the active site of amylases[10]. Oligomeric substrates can interact productively or non-productively with the subsites. The productive binding mode means that a susceptible bond lies over the catalytic site, in which case the bond is cleaved. Endo-acting enzymes form several productive binding complexes resulting in a given product pattern. The relative rate of formation of each product is called bond cleavage frequency (BCF), which can be used to determine the subsite-binding energies. Computer program SUMA (Subsite Mapping of α-Amylases) developed by Gyémánt and co-workers[11] uses BCFs of a series of oligomeric substrates to calculate the position of the cleavage site, the number of subsites and the binding energy of each subsite except of the two central subsites (subsites −1 and +1) adjacent to the catalytic site, which are occupied in all productive complexes. The distribution of the high and low affinity binding subsites and the barrier subsites within the active site determines the hydrolytic efficiency on series of substrates with various length: high-affinity subsites next to the cleavage sites (−2 through +2) allow effective cleavage of shorter substrates, while distal high-affinity subsites control the action on longer substrates[12]. Chain length and concentration of characteristic products can be determined from the BCF tables and energy contribution of each subsite to the total binding energy can be calculated. However, experimental procedure of BCF is a time-consuming and expensive procedure because it requires labeled oligosaccharide (for example: CNP-MOS) series as substrate and separation and quantification of each product by an HPLC-method[10].

Computer-aided analysis of X-ray crystallographic structures is an effective way to examine enzyme-substrate interactions and it could be supplemented by molecular mechanical/dynamical calculations to predict the interaction in a quantitative manner. In this work, we aimed to develop a computer-aided procedure for subsite mapping of α-amylases to predict and interpret functional features of wild type and mutant α-amylases.

2. Materials and methods

2.1. Calculation of subsite binding energies and BCFs by SUMA

Experimentally determined BCFs were collected from the literature for each enzyme and based on them, number of subsites, position of the cleavage site and binding energy (E_{SUMA}) of each subsite were calculated by SUMA. Apparent free energy values were optimized in a SUMA calculation by minimization of differences between experimental and calculated BCFs[11]. Differences of the calculated subsite binding energies were below 0.4 kJ/mol in case of two independent experimental measurements of action pattern of the same enzyme (data not shown). To standardize data evaluation, E_{SUMA} and E_{Sybyl} data (see later) were calculated for the same number of subsites of wild-type and mutant enzymes. The temperature and the interpolation step were set to 37 °C and 0.01, respectively, in all calculation in contrast to the default values of the program.

BCF prediction using calculated subsite binding energies was performed by SUMA according to the standard parameters (substrates of DP 3 to 11, temperature 37 °C).

2.2. Molecular modeling

Crystal structures of wild-type enzymes were downloaded from the Protein Data Bank[13]. Structures of human salivary α-amylase (PDB code: 1SMD[14]), barley α-amylase AMY1 (PDB code: 1RP8[15]) and AMY2 (PDB code: 1BG9[16]), α-amylase of *Bacillus amyloliquefaciens* (PDB code: 3BH4[17]), porcine pancreatic α-amylase isoenzyme II (PPA II) (PDB code: 1PIG[18]) and α-amylase of *Aspergillus oryzae* (PDB code: 7TAA[19]) were used as templates for building of 3D structures of wild type and mutant enzymes using Sybyl program package (Tripos Inc., St. Louis, MO, USA). Structures of the maltooligosaccharide substrates were modeled based on the crystal structures of maltose (PDB code: 2GVY[20]), maltoheptaose (PDB code: 1RP8[15]) and a substrate-analogue inhibitor acarbose (PDB code: 1MFU[21], 1RPK[15], 1E3Z[22], 1OSE[23]) complexed with various alpha-amylases.

Initial structure of an enzyme-substrate complex was generated by merging the substrate into the binding cleft of the enzyme and deleting the inhibitor if it existed in the original crystal structure. Water molecules bumped with substrate were also removed. In mutant proteins, wild-type residues were changed by Sybyl. To avoid bumping of mutant residues with unmodified ones, water molecules or substrate chain, the spatial position of the side chain of the modified residue was predicted by a short molecular dynamics run (length: 1000 step, step size: 1 fs, temperature: 300 K, dielectric constant: 1, AMBER7_FF99 force field). To refine the position of the substrate and the mutated residues, a short minimization procedure was performed by Sybyl (100 Powell iterations, dielectric constant 1, AMBER7_FF99 force field) on the 8 Å surround of the oligosaccharide chain and the mutated residues. The resulted complex was further energy-minimized without any fixed atoms by Sybyl using the following parameters: AMBER7_FF99 force field, non-bonded cutoff 8 Å, 100 Powell iterations and dielectric constant 1. Interaction energy (E_{Sybyl}) between the enzyme and the carbohydrate substrate was calculated for each subsite. Calculations and visualization were performed on Silicon Graphics Fuel workstations (Silicon Graphics International, Fremont, CA, USA).

3. Results and discussion

3.1. Data sets

The training set consisted of the subsite binding energies of all wild-type enzymes and some mutants of human salivary amylase (HSA) and barley amylase 1 (AMY1) enzymes. Wild-type enzymes were chosen based on the existence of experimental BCF values as well as available crystal structure. In contrast to the various degree of sequence identity (11-86 %) between the wild-type enzymes (Table 1), they were all classified into the glycoside hydrolase family 13 and showed high degree of structural conservation. Subsite binding energy values of HSA enzymes were recalculated by SUMA from the previously published action patterns of wild type HSA and its W58L[24] and Y151M[25] mutants. Recalculation of published subsite maps was necessary, because the published energies were calculated for 8 subsites (5 glycone and 3 aglycone subsites); however, reliable structural model for substrate can be built only for 7 subsites (4 glycone and 3 aglycone subsites). The published binding energy of subsite −5 was small (0.65 kJ/mol)[24], therefore simplification of the subsite model should cause only marginal changes in the binding energies of other subsites. Indeed, the original and the recalculated binding energies were in excellent agreement, square of the correlation coefficients (r^2) from the linear regression analysis were 0.986. Similarly, published subsite maps of all other wild-type

enzymes were recalculated and there were excellent correlations in all cases (Table 2). The training set consisted of total 13 wild-type and mutant enzymes and 97 subsite binding energy values (Table 3).

A test set was created from subsite binding energies of novel mutants of AMY1: S48Y, V47A, V47D, V47F, V47I/S48I, V47K/S48G, V47G/S48D and V47L/S48A. These single and double mutant enzymes contain various mutations of V47 and/or S48 residues which were engineered previously with random mutagenesis for the investigation of the enzyme-substrate interactions[26]. Action patterns of these enzymes were determined experimentally using hydrolysis of CNP-MOSs of DP 3−11 and subsite maps were calculated by SUMA (Mori et al., in preparation). The subsite maps were calculated for 7 glycone and 4 aglycone binding sites by the same way as the AMY1 group of training set. The test set consisted of total 8 mutant enzymes and 73 subsite binding energy values (Table 3).

3.2. Steps of calculation

Our working hypothesis was that the subsite binding energies could be predicted from enzyme-substrate interaction energies calculated by a molecular mechanical program. Homologous models of enzyme-substrate complexes of the training set were built and interaction energies were calculated between the protein and the carbohydrate residues of each subsite (E_{Sybyl}). These values were compared to those calculated based on the experimentally determined action patterns (E_{SUMA}). To reach highest correlation between E_{SUMA} and E_{Sybyl} data, parameters of the energy-minimization procedure were optimized. Several force fields available in Sybyl program: Kollman_All_Atom, Amber95 and Amber7_FF99 were tested, together with an in-house implementation of Glycam06 force field[27]. Amber7_FF99 was chosen because its performance was the best in our hand. Furthermore, wide-variety of values of dielectric constant (1, 2, 3, 4 and 8), non-bonded cutoff (8, 10 and 12), number of iterations (20-100 by steps of 20 and 100-1000 by steps of 100) and the width of minimized shell (1-12 angstrom) of atoms around the substrate and the mutated residues in the initial energy minimization were also probed to find a good combination of parameters. Our aim was to build a relatively fast procedure, therefore explicit molecules for bulk solvent and extensive minimization were avoided. Linear regression analysis was performed with the exception of energies of the −1 and +1 subsites because it is theoretically impossible to calculate them from experimental bond cleavage frequency data[11]. After regression analysis, E_{Sybyl} data were linearly transformed to the scale of E_{SUMA} data with the values of intercept and slope of the best fit line to get $E_{Transf.}$ (Fig. 2A and 2B) for each enzyme group. Enzyme groups were based on the wild-type enzymes and

each mutant belonged to the group of the corresponding wild-type enzyme. Calculated E_{Sybyl} and $E_{Transf.}$ values of training enzymes (Table 3) showed good correlation with E_{SUMA} values: r^2 were in the range of 0.827−0.929 (Table 4).

The procedure was applied for the test set to assess the predictive potential of the model. Calculated interaction energies (E_{Sybyl}) of the test set (containing only AMY1 mutants) were transformed to $E_{Transf.}$ values using equation of AMY1 group of the training set. Calculated E_{Sybyl} and $E_{Transf.}$ values of the test set (Table 3) showed correlation with E_{SUMA} values (Fig. 2C): r^2 was 0.502 (Table 4) which was lower than the corresponding value of 0.827 of AMY1 group of the training set, as expected. However, the error distribution was odd: all values out of 95% prediction interval were belonged to V47D, V47K/S48G and V47G/S48D mutants. All these mutants and only these mutants contained a substitution from neutral to charged residue. If all values of these mutants were omitted from the plot, the r^2 value rose up to 0.638.

Mutants of the test set contained single or double mutations in the position of V47 and/or S48 residues of AMY1. Crystallographic analysis showed that Val47 was involved in direct hydrogen bonds at the −7 subsite, while Ser48 formed indirect hydrogen bonds at the −1 and −2 subsites[15]. Docking studies by molecular modeling predicted that the influence of V47 was extended to the −6, −5, −4 subsites and S48 effected substrate binding at the −4, −3, −2 subsites[28]. Furthermore, V47 together with Y105, may be considered as the substrate "entrance" to the active site cleft where the substrate adopts a half-circle conformation centered on Val47[15]. Therefore adjacent V47 and S48 amino acids are surrounded by several glucose residues of bounded substrate chain and they may effect the dynamics of enzyme-substrate interaction in addition to influence of binding affinity. It seems that our simple procedure was only moderately able to model this combined effect as the correlation coefficient was dropped from 0.827 to 0.638 (neglecting mutants having substitutions to charged residues). Large changes in charge distribution of the mutated position turned out to be also problematic, as in cases of V47D, V47K/S48G and V47G/S48D mutants: the correlation coefficient was further dropped to 0.502. In contrast with the test set, enzymes of AMY1 group of training set contain the mutations of Y105 and/or T212, which amino acids located at −6 and +4 subsites and participated in substrate binding at only one or two outermost substrate binding site[28] and these mutants did not harbor any substitution to charged residues.

Some mutants containing amino acid changes in the secondary binding sites were also tested, but our procedure failed to correctly predict changes caused by these distant sites, as expected (data not shown).

3.3. Prediction of action pattern

Computer program SUMA is capable not only to calculate E_{SUMA} energies but also to recalculate BCFs from subsite maps[11] thus it can be used for prediction of the primary experimental values. Bond cleavage frequencies were calculated using the $E_{Transf.}$ values as input data in SUMA and the predicted BCFs were correlated with the experimentally determined values. Linear regression analysis resulted in good correlation (r^2 = 0.727−0.835) between predicted and published BCFs of enzymes of the training set (Table 4). In case of mutant AMY1 enzymes of the test set, the correlation coefficient was similarly low to that of binding energy analysis: r^2 = 0.538 (Table 4). Not unexpectedly, the odd error distribution was also present here: if values of mutant enzymes having substitution to charged residues (V47D, V47K/S47G and V47G/S48D) were omitted, the value of r^2 rose up to 0.728.

4. Conclusion

The main structural feature of family 13 glycoside hydrolases is the presence of the characteristic $(\beta/\alpha)_8$-barrel domain with a varying number of extra domains. The substrate binding site is made from the residues of domains A and B, however, the architecture of the domain B varies between amylases and it is the major determinant in differences of substrate specificities. Present study was made with the aim to develop a computer-aided subsite mapping procedure to better understand the specificity of α-amylases. Our procedure was successfully adopted for α-amylases belonging to different kingdoms: BAA represents liquefying bacterial α-amylases from Bacteria, further α-amylases were derived from Eukaryota: TAA represents amylases from fungi, AMY1 and AMY2 represent plant enzymes and PPA and HSA represent the mammalian enzymes. The structural conservation of these enzymes were much stronger than the sequence conservation (Table 4), which give the hope that our procedure can be successfully applied for wide range of α-amylases.

The great clinical and industrial importances justify the studies of α-amylases. Described procedure could help to generate subsite maps of α-amylases as well as it could predict the primary experimental data (BCFs) via the use of computer program SUMA[11]. However, special care is needed when neutral residues was changed to charged one (and probably *vice versa*) in the substrate binding site of an α-amylase. The computer-aided subsite mapping may support the examination of the role of substrate binding residues and it can be used to supply protein engineers in design of modified enzymes with altered action pattern. Our method can

supplement the experimental subsite mapping procedure by prediction of subsite maps and action patterns in a fast and cheap way.

Acknowledgements

Erika Fazekas is acknowledged for help in determination of BCFs.

References

1. Gupta, R.; Gigras, P.; Mohapatra, H.; Goswami, V.K.; Chauhan, B. *Process Biochem.* **2003**, *38*, 1599-1616.
2. Rohleder, N.; Nater, U.M. *Psychoneuroendocrinology* **2009**, *34*, 469-485.
3. Cantarel, B.L.; Coutinho, P.M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. *Nucleic Acids Res.* **2009**, *37*, D233-D238.
4. MacGregor, E.A.; Janecek, S.; Svensson, B. *Biochim. Biophys. Acta* **2001**, *1546*, 1-20.
5. Janecek, S.; Svensson, B.; Henrissat, B. *J. Mol. Evol.* **1997**, *45*, 322-331.
6. Jespersen, H.M.; Macgregor, E.A.; Sierks, M.R.; Svensson, B. *Biochem. J.* **1991**, *280*, 51-55.
7. Janecek, S.; Svensson, B.; MacGregor, E.A. *Eur. J. Biochem.* **2003**, *270*, 635-645.
8. Kumar, V. *Carbohydr. Res.* **2010**, *345*, 893-898.
9. Davies, G.J.; Wilson, K.S.; Henrissat, B. *Biochem. J.* **1997**, *321*, 557-559.
10. Kandra, L.; Gyémánt, G.; Lipták, A. *Biologia* **2002**, *57*, 171-180.
11. Gyémánt, G.; Hovánszki, G.; Kandra, L. *Eur. J. Biochem.* **2002**, *269*, 5157-5162.
12. Kandra, L.; Abou Hachem, M.; Gyémánt, G.; Kramhoft, B.; Svensson, B. *FEBS Lett.* **2006**, *580*, 5049-5053.
13. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. *Nucleic Acids Res.* **2000**, *28*, 235-242.
14. Ramasubbu, N.; Paloth, V.; Luo, Y.G.; Brayer, G.D.; Levine, M.J. *Acta Crystallogr. D Biol. Crystallogr.* **1996**, *52*, 435-446.
15. Robert, X.; Haser, R.; Mori, H.; Svensson, B.; Aghajari, N. *J. Biol. Chem.* **2005**, *280*, 32968-32978.
16. Kadziola, A.; Sogaard, M.; Svensson, B.; Haser, R. *J. Mol. Biol.* **1998**, *278*, 205-217.

17. Alikhajeh, J.; Khajeh, K.; Ranjbar, B.; Naderi-Manesh, H.; Lin, Y.H.; Liu, E.H.; Guan, H.H.; Hsieh, Y.C.; Chuankhayan, P.; Huang, Y.C.; Jeyaraman, J.; Liu, M.Y.; Chen, C.J. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2010**, *66*, 121-129.
18. Machius, M.; Vertesy, L.; Huber, R.; Wiegand, G. *J. Mol. Biol.* **1996**, *260*, 409-421.
19. Brzozowski, A.M.; Davies, G.J. *Biochemistry* **1997**, *36*, 10837-10845.
20. Vujicic-Zagar, A.; Dijkstra, B.W. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2006**, *62*, 716-721.
21. Ramasubbu, N.; Ragunath, C.; Mishra, P.J. *J. Mol. Biol.* **2003**, *325*, 1061-1076.
22. Brzozowski, A.M.; Lawson, D.M.; Turkenburg, J.P.; Bisgaard-Frantzen, H.; Svendsen, A.; Borchert, T.V.; Dauter, Z.; Wilson, K.S.; Davies, G.J. *Biochemistry* **2000**, *39*, 9099-9107.
23. Gilles, C.; Astier, J.P.; MarchisMouren, G.; Cambillau, C.; Payan, F. *Eur. J. Biochem.* **1996**, *238*, 561-569.
24. Ramasubbu, N.; Ragunath, C.; Mishra, P.J.; Thomas, L.M.; Gyémánt, G.; Kandra, L. *Eur. J. Biochem.* **2004**, *271*, 2517-2529.
25. Kandra, L.; Gyémánt, G.; Remenyik, J.; Ragunath, C.; Ramasubbu, N. *FEBS Lett.* **2003**, *544*, 194-198.
26. Svensson, B.; Jensen, M.T.; Mori, H.; Bak-Jensen, K.S.; Bonsager, B.; Nielsen, P.K.; Kramhoft, B.; Praetorius-Ibba, M.; Nohr, J.; Juge, N.; Greffe, L.; Williamson, G.; Driguez, H. *Biologia* **2002**, *57*, 5-19.
27. Kirschner, K.N.; Yongye, A.B.; Tschampel, S.M.; Daniels, C.R.; Foley, B.L.; Woods, R.J.J. *J. Comput. Chem.* **2008**, *29*, 622-655.
28. Bak-Jensen, K.S.; Andre, G.; Gottschalk, T.E.; Paes, G.; Tran, V.; Svensson, B. *J. Biol. Chem.* **2004**, *279*, 10093-10102.
29. Allen, J.D.; Thoma, J.A. *Biochem. J.* **1976**, *159*, 121-131.

FIGURE LEGENDS

Figure 1. Schematic representation of structure of AMY1 (bottom panel) together with the architecture of substrate binding site of studied wild-type enzymes (top panel). Green hexagons represent monosaccharide units of the substrate, large arrow shows the site of cleavage, subsites are labeled by negative and positive numbers for glycone and aglycone binding sites, respectively.

Figure 2. Correlation between the E_{SUMA} values of AMY1 group and the predicted energies (E_{Sybyl} and $E_{Transf.}$). Black lines correspond to the best fitted regression line and green lines show the 95 % prediction interval range.

A) AMY1 group of training set **before** linear transformation ($E_{Sybyl} = 0.905 \times E_{SUMA} - 17.731$)

B) AMY1 group of training set **after** linear transformation ($E_{Transf} = 1.000 \times E_{SUMA} - 0.000$)

C) AMY1 group of **test set** after linear transformation.

Table 1. Sequence identities (%) of the studied enzymes based on structural alignment (UniProtKB codes of protein sequences used for sequence alignment: HSA – P04745, PPA – P00690, AMY1 – P00693, AMY2 – P04063, TAA – P0C1B3, BAA – P00692).

	HSA	**PPA**	**AMY1**	**AMY2**	**TAA**	**BAA**
HSA	100	--	--	--	--	--
PPA	86	100	--	--	--	--
AMY1	12	11	100	--	--	--
AMY2	12	11	74	100	--	--
TAA	18	19	14	13	100	--
BAA	16	18	15	15	20	100

Table 2. Original and modified subsite models of the studied enzymes and the correlation between the subsite binding energies of the original publications and the recalculated values based on subsite models of this work.

Enzyme	Subsite model		square of correlation coefficient (r^2)	reference
	original work	this work		
HSA	−5 ... +3	−4 ... +3	0.986	[24, 25]
PPA	−4 ... +3	−4 ... +3	0.942	[11]
AMY1	−8 ... +4	−7 ... +4	0.994	[12]
AMY2	−8 ... +4	−7 ... +4	0.999	[12]
TAA	−3 ... +5	−3 ... +5	-- *	[29]
BAA	−6 ... +4	−7 ... +3	0.996	[29]

* Energy values of subsites were not published in the original work.

Table 3. Calculated subsite binding energies (kJ/mol) of studied enzymes.

Enzyme		Energy	Subsite												Set
			−7	−6	−5	−4	−3	−2	−1	+1	+2	+3	+4	+5	
PPA	**wt**	E_{SUMA}				-0.86	-14.00	-19.99			-15.48	-0.04			**Training**
		$E_{Transf.}$				-1.66	-10.66	-19.54			-19.03	0.53			
AMY2	**wt**	E_{SUMA}	-3.21	-8.39	1.31	-2.07	1.51	-5.69			-3.53	1.33	0.25		
		$E_{Transf.}$	-3.00	-8.57	-1.23	-0.37	3.27	-4.80			-5.84	1.41	0.66		
BAA	**wt**	E_{SUMA}	-3.99	-8.13	-13.21	-4.12	-7.57	-13.15			-10.68	-7.48			
		$E_{Transf.}$	-1.29	-7.68	-13.56	-7.08	-8.05	-12.54			-11.37	-6.76			
TAA	**wt**	E_{SUMA}					-1.62	-19.99			-10.57	-4.9	2.26	-3.36	
		$E_{Transf.}$					0.43	-20.79			-8.14	-4.19	2.59	-8.09	
HSA	**wt**	E_{SUMA}				-1.90	-6.74	-9.04			-12.25	-1.72			
		$E_{Transf.}$				-0.77	-2.74	-7.75			-7.10	-0.68			
	W58L	E_{SUMA}				-0.08	-2.10	-5.72			-6.15	-0.81			
		$E_{Transf.}$				-0.74	-3.26	-7.14			-6.29	-0.14			
	Y151M	E_{SUMA}				-1.56	-6.51	-6.03			-2.99	-0.11			
		$E_{Transf.}$				2.36	-5.54	-7.45			-3.82	0.42			
AMY1	**wt**	E_{SUMA}	-4.24	-12.47	0.65	-1.30	0.92	-7.23			-2.86	1.99	0.59		
		$E_{Transf.}$	-3.63	-9.53	-0.47	-3.09	1.76	-8.56			-5.12	2.58	0.77		
	Y105A	E_{SUMA}	-4.06	-5.89	-0.11	-1.63	-0.05	-8.89			-5.27	1.72	0.43		
		$E_{Transf.}$	-3.30	-5.09	2.92	-3.65	1.69	-8.08			-5.88	2.50	0.65		
	Y105F	E_{SUMA}	-4.06	-13.59	0.75	-2.19	-0.25	-7.72			-4.07	0.91	0.01		
		$E_{Transf.}$	-3.24	-9.29	0.16	-3.15	1.69	-8.36			-5.84	1.96	0.91		
	Y105W	E_{SUMA}	-4.38	-10.84	0.74	-1.44	0.32	-7.06			-3.08	1.33	0.08		
		$E_{Transf.}$	-3.95	-10.85	0.84	-2.00	1.69	-8.12			-5.63	2.73	0.68		
	T212Y	E_{SUMA}	-4.05	-10.06	-1.40	-2.31	0.22	-4.72			-4.24	0.48	-0.20		
		$E_{Transf.}$	-3.50	-9.44	-0.31	-2.24	1.58	-8.67			-8.54	1.72	-1.88		
	Y105A/ T212Y	E_{SUMA}	-3.22	-5.49	-0.82	-1.70	0.75	-4.37			-6.85	1.94	0.00		
		$E_{Transf.}$	-3.41	-5.26	2.40	-3.23	1.90	-8.49			-8.64	1.42	-0.34		
	S48Y	E_{SUMA}	-2.41	-1.07	0.4	1.37	0.77	-8.16			-7.32	2.36	0.09		**Test**
		$E_{Transf.}$	-4.50	-7.56	-0.82	-1.28	3.06	-10.71			-6.96	1.67	-0.62		
	V47A	E_{SUMA}	-3.07	-9.49	1.98	-1	1.2	-5.05			-2.72	2.87	0.9		
		$E_{Transf.}$	2.62	-6.03	0.90	1.16	1.27	-9.34			-5.37	2.26	0.70		
	V47F	E_{SUMA}	-7.32	-4.44	0.83	-1	0.08	-9.05			-5.85	2.4	1.32		
		$E_{Transf.}$	-3.69	-4.33	-2.01	-4.91	1.78	-9.44			-5.23	2.76	0.69		
	V47D	E_{SUMA}	0.15	-7.54	-0.25	-1.86	-0.91	-8.88			-5.97	1.45	-0.35		
		$E_{Transf.}$	9.71	-13.42	-9.53	-4.95	2.13	-10.52			-5.47	2.09	0.80		
	V47K/ S48G	E_{SUMA}	-2.22	-1.29	0.8	1.84	-0.91	-10.81			-8.02	1.92	0.2		
		$E_{Transf.}$	-11.95	1.80	-7.00	1.71	3.18	-8.79			-5.17	2.47	0.73		
	V47G/ S48D	E_{SUMA}	-3.44	-2.53	1.63	-0.69	-0.76	-8.52			-5.93	1.76	-0.53		
		$E_{Transf.}$	4.19	-7.25	1.24	1.12	7.64	-12.11			-5.44	2.31	0.81		
	V47I/ S48I	E_{SUMA}	-3.98	-10.32	-0.48	0.88	-1.07	-5.58			-6.43	2.22	1.54		
		$E_{Transf.}$	-2.13	-6.01	-0.14	-1.78	3.05	-8.10			-5.60	2.51	0.37		
	V47L/ S48A	E_{SUMA}	-4.24	-8.77	0.06	-2.11	0.13	-5.94			-5.07	1.82	0.64		
		$E_{Transf.}$	-1.82	-5.95	-1.57	-3.06	2.72	-8.44			-4.91	2.23	0.72		

Table 4. Parameters of the best fitted line of linear regression analysis of E_{SUMA} and E_{Sybyl} values (middle panel) and the experimental and calculated values of BCFs (right panel) (*m* − slope, *b* − intercept, r^2 − square of correlation coefficient).

Enzyme		Subsite binding energy			Action pattern			Set
		m	b	r^2	m	b	r^2	
PPA	**wt**	0.510	-16.261	0.929	1.070	-2.054	0.835	**Training**
AMY2	**wt**	1.133	-17.106	0.839	0.912	1.616	0.756	
BAA	**wt**	1.044	-12.884	0.834	0.789	5.224	0.732	
TAA	**wt**	1.188	-13.403	0.902	0.988	0.171	0.782	
HSA	**wt**	0.912	-15.999	0.888	0.821	3.587	0.727	
	W58L							
	Y151M							
AMY1	**wt**	0.905	-17.731	0.827	0.936	1.229	0.801	
	Y105A							
	Y105F							
	Y105W							
	T212Y							
	Y105A/T212Y							
	S48Y	0.844	-17.888	0.502	0.838	3.042	0.538	**Test**
	V47A							
	V47F							
	V47D							
	V47K/S48G							
	V47G/S48D							
	V47I/S48I							
	V47L/S48A							

Research highlights

- Subsite binding energies of homologous models of wild-type and mutant α-amylases were calculated by a molecular mechanical program.

- Parameters of the calculation were set up to get correlation between the calculated and the experimental binding energies of enzymes of a training set.

- Calculations on an independent test set also resulted in good correlation, except mutants containing neutral to charged residue substitution.

- Our computer-aided procedure may help to understand structure-function relationship of α-amylases and to design enzymes with new features.

Figure1

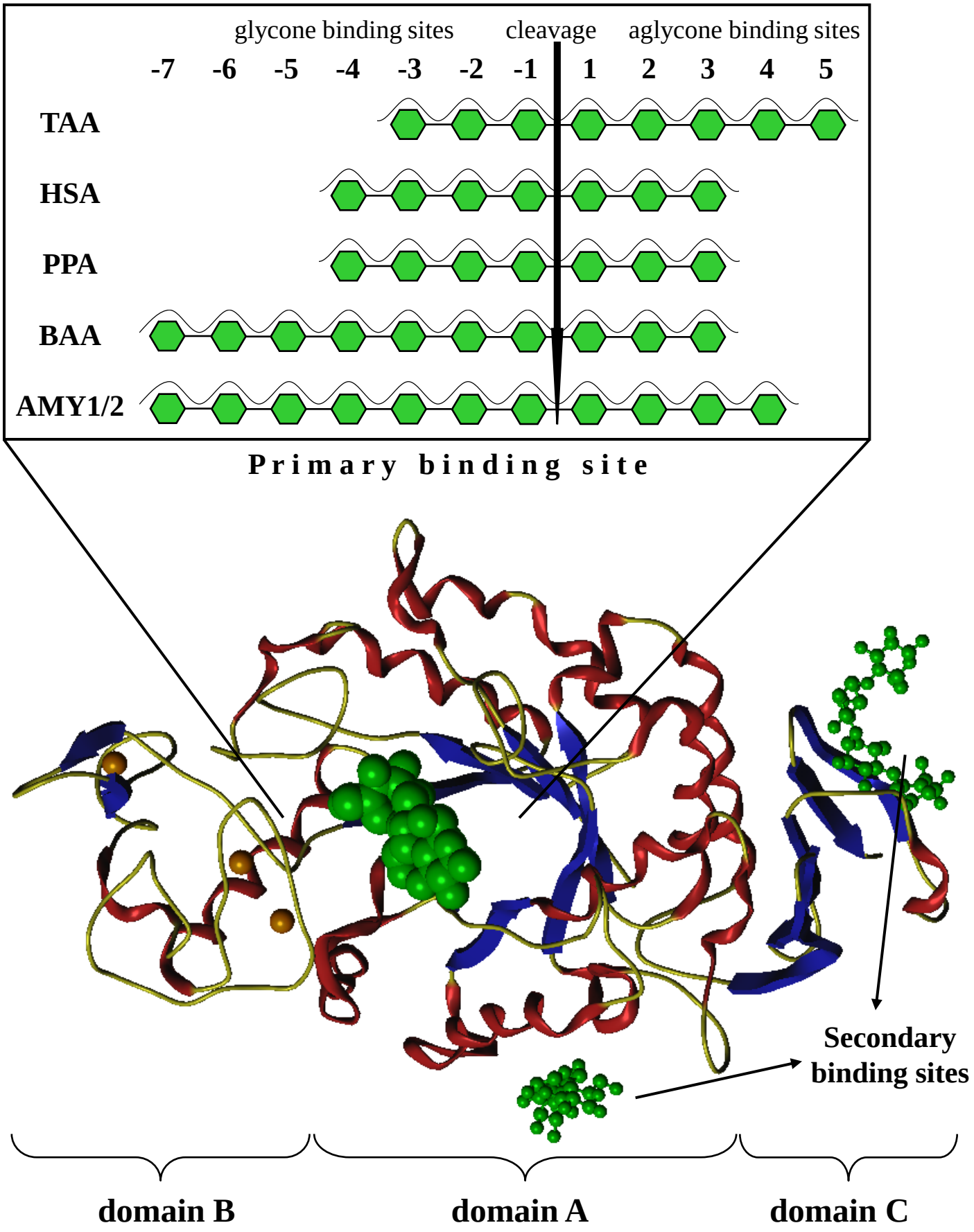

glycone binding sites
cleavage
aglycone binding sites
-7 -6 -5 -4 -3 -2 -1 1 2 3 4 5
TAA
HSA
PPA
BAA
AMY1/2
Primary binding site
Secondary binding sites
domain B
domain A
domain C

Figure2A

Click here to download high resolution image

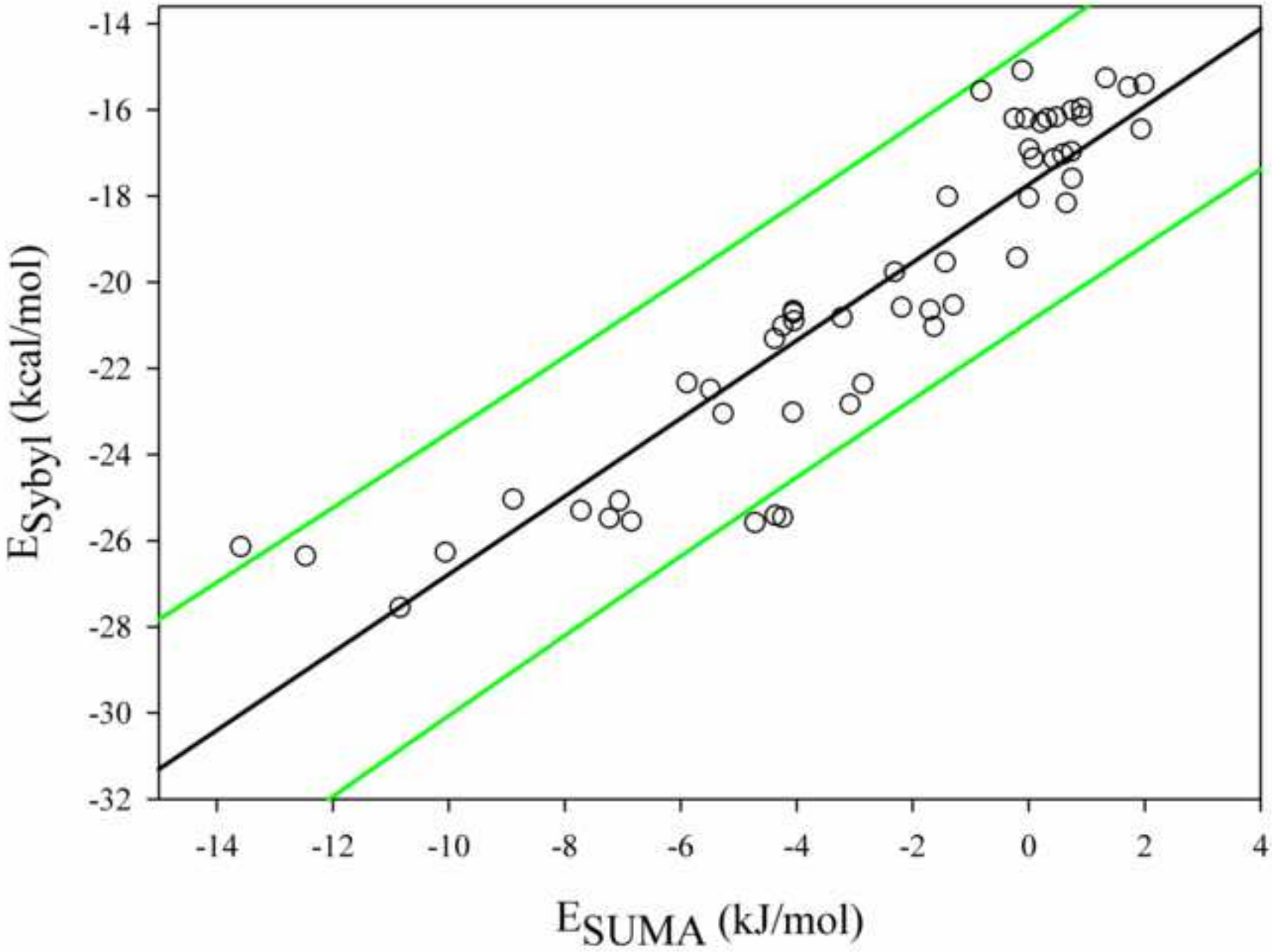


Figure2B

Click here to download high resolution image

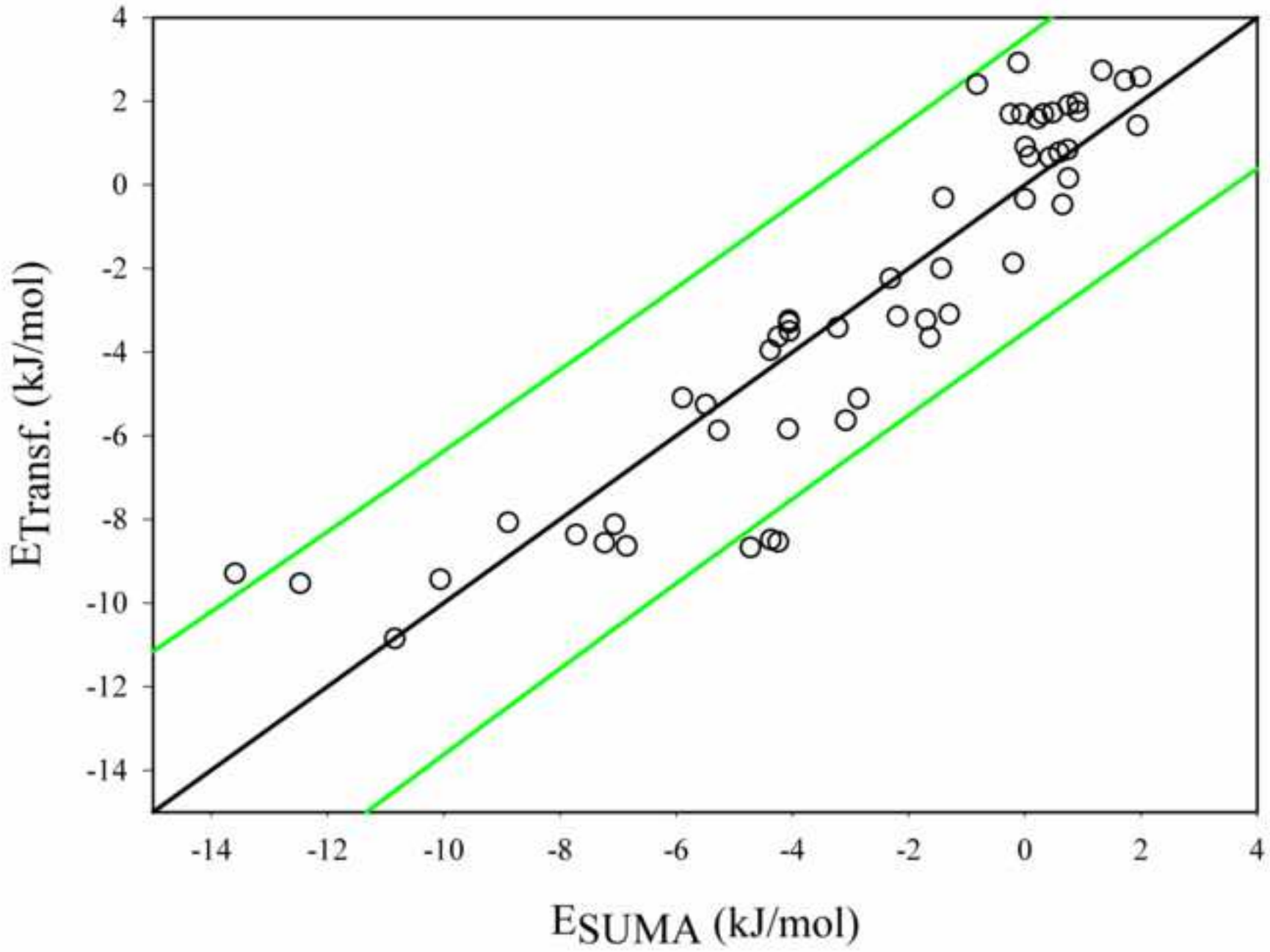


Figure2C

Click here to download high resolution image