Published online 21 May 2021 *Nucleic Acids Research, 2021, Vol. 49, Web Server issue* ***W15–W20***
https://doi.org/10.1093/nar/gkab377

Amino Acid Interactions (INTAA) web server v2.0: a single service for computation of energetics and conservation in biomolecular 3D structures

Jiří Vymětal[1,†], David Jakubec[1,2,*,†], Jakub Galgonek[1] and Jiří Vondrášek[1,*]

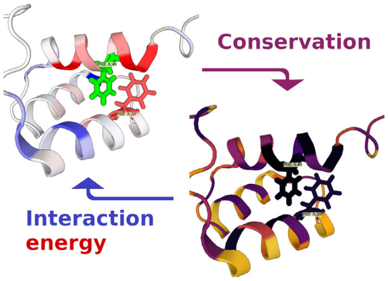
[1]Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Praha 6, 160 00, Czech Republic and [2]Department of Software Engineering, Faculty of Mathematics and Physics, Charles University, Praha 1, 118 00, Czech Republic

Received March 10, 2021; Revised April 10, 2021; Editorial Decision April 24, 2021; Accepted April 27, 2021

ABSTRACT

Interactions among amino acid residues are the principal contributor to the stability of the three-dimensional structure of a protein. The Amino Acid Interactions (INTAA) web server (https://bioinfo.uochb.cas.cz/INTAA/) has established itself as a unique computational resource, which enables users to calculate the contribution of individual residues in a biomolecular structure to its total energy using a molecular mechanical scoring function. In this update, we describe major additions to the web server which help solidify its position as a robust, comprehensive resource for biomolecular structure analysis. Importantly, a new continuum solvation model was introduced, allowing more accurate representation of electrostatic interactions in aqueous media. In addition, a low-overhead pipeline for the estimation of evolutionary conservation in protein chains has been added. New visualization options were introduced as well, allowing users to easily switch between and interrelate the energetic and evolutionary views of the investigated structures.

GRAPHICAL ABSTRACT



INTRODUCTION

The stable arrangement of amino acid residues known as the three-dimensional (3D) structure of a protein is the result of various covalent and noncovalent interactions among the residues as well as interactions with the environment, i.e. the molecules of the solvent. The overall stability of a protein can be quantified as the Gibbs free energy difference between the folded and unfolded states. It has been experimentally shown that, in many proteins, only a fraction of residues is essential for maintaining their 3D structure and function while the rest can be mutated without any major effect on stability (1).

Amino acid residues critical for protein stability can be identified and analyzed using, for example, the theoretical methods of computational chemistry and biophysics. These residues typically participate in noncovalent interactions characterized by highly favorable interaction energies (IEs). The IEs of individual amino acid residue pairs can be arranged in the form of a square interaction energy matrix (IEM), which has been shown to be a powerful tool for the identification of structurally and functionally significant residues (1,2).

In addition to the physical, energetic view of the biomolecular structure, evolutionary conservation can be an important indicator of the residues' structural and functional significance. For example, amino acid residues involved in hydrophobic core formation or enzymatic activity usually display lower substitution rates compared to other sites on the protein (3). Evolutionary conservation, quantified using, e.g. the information content (IC) of individual columns in a multiple sequence alignment (MSA) of related sequences, can thus provide information orthogonal to the analysis of IEs. IC-based mark-up can be obtained for protein 3D structures using web applications such as ConSurf (3) and 3DPatch (4); however, bridging the energetic analy-

[*]To whom correspondence should be addressed. Tel: +420 220 183 294; Email: david.jakubec@uochb.cas.cz
Correspondence may also be addressed to Jiří Vondrášek. Tel: +420 220 183 267; Email: jiri.vondrasek@uochb.cas.cz.
Present address: David Jakubec, Department of Software Engineering, Faculty of Mathematics and Physics, Charles University, Praha 1, 118 00, Czech Republic.
[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

ses and the projections of conservation has so far not been possible to do using a single web server.

The Amino Acid Interactions (INTAA) web server (https://bioinfo.uochb.cas.cz/INTAA/) is a unique computational resource which focuses on the investigation of energetics of interactions among the constituting moieties (e.g. amino acid residues) of a biomolecular 3D structure (5). The web server's functionality is split into two services: the IEM web application, and the amino acid–nucleotide interactions web server. The latter of these components provides a static analysis of the geometric and energetic preferences in the interactions between amino acids and nucleotides in a predefined set of protein–DNA complexes. As this component has not received any major updates in this release of the INTAA web server, we will further focus on the IEM web application and refer to it interchangeably using the INTAA acronym.

The IEM web service is an interactive application which accepts as its input a biomolecular 3D structure [specified using a Protein Data Bank (PDB) (6) identifier (ID) or uploaded as a file by the user] and calculates the pairwise IEs among its constituting moieties. As described below, the query structure is automatically processed by the web server, hydrogen atoms are optionally added, and the IEs are calculated using a molecular mechanical force field (FF). The user can further choose a model used to simulate the environment as a dielectric continuum. The IEM web service then calculates for each residue in the query structure the pairwise IEs for its interactions with all other residues in the structure, as well as its total (net) IE, defined as the sum of the former. The pairwise and total IEs can be viewed numerically in the form of interactive tables, as well as shown projected onto the query structure in the integrated molecular structure viewer.

The INTAA web server has been used recently, for example, for the investigation of effective pairwise potentials between amino acid residues in proteins (7), for the comparison of the strength of interactions of particular amino acid residues (8), or as a reference method for the calculation of IEs (9). The INTAA web server has been further employed in the design of peptidic cysteine protease inhibitors targeted against the malaria parasite (10), for the elucidation of the structure–function relationship in a cyanobacterial Mn-catalase enzyme (11), and for the investigation of a missense mutation in a human UDP-glucose dehydrogenase linked with developmental delay (12).

In this update, we describe the major functions added to the INTAA web server since its original release. Notably, a new continuum solvation model has been introduced as the default option for the treatment of electrostatic interactions in aqueous media, providing a more accurate representation of the interactions. A new pipeline for the calculation of evolutionary conservation of amino acid residues in the protein chains has been developed and added to the web server. The pipeline is completely automatic and is invoked with each new task. New selection and visualization options have been further added, allowing the users to easily interrelate the energetic and evolutionary views of the query structures. These additions help establish the INTAA web server and the IEM web service as comprehensive tools for the interactive analysis of biomolecular 3D structures. In the following sections, technical description of the new features and examples of their usage are provided.

MATERIALS AND METHODS

New continuum solvation model

The IEM web service calculates the IE as a sum of the noncovalent van der Waals and point-charge electrostatic interactions. The electrostatic term can be evaluated using Coulomb's law, neglecting any environmental effects, or the interaction with the solvent can be introduced in an implicit manner. In the previous version of the application, the implicit solvent was modeled using a distance-dependent dielectric constant (DDDC). Although such treatment of the solvent decreases the magnitude of the electrostatic term at large distances, it was an *ad hoc* approach without solid physical grounding. Simplicity, ease of implementation, and zero performance overhead were the only advantages of this approach. On the other hand, the Poisson–Boltzmann (PB) theory (13) provides proper description of continuum electrostatics. However, solving the corresponding electrostatic field equations requires numerical integration over discretized space elements, making it time-consuming and sensitive to implementation details. Furthermore, the nonadditive character of the PB theory does not allow a straightforward separation of the pairwise interaction terms which are the cornerstone of the IEM method.

In order to improve the quality and accuracy of the electrostatic term evaluated in the dielectric continuum, the generalized Born (GB) model was adopted in the current version of the application. We have shown in the past that empirical potentials along with the GB model are able to reproduce the results of quantum chemical DFT-D/PCM calculations while being fast enough to be used for calculations involving large biological systems (14). The GB theory is the limit case of the PB theory obtained under several approximations (15). Near complete match between the GB and PB theories can be achieved when the Born radii—the critical parameters of the GB model—are calculated exactly (16). The Born radius of an atom expresses its effective distance from the dielectric boundary and hence its exposure to the solvent. Nevertheless, estimation of the Born radii is the most approximative and heuristic part of the GB theory applications. The algorithm newly implemented in the IEM web service follows the approach of Onufriev, Bashford and Case, namely the parametrization II (OBC II) described in (17). As the GB model takes the solvent exposure of the atoms into account explicitly, it clearly surpasses the accuracy of models which utilize a DDDC. In addition, the GB model permits the pairwise treatment of the electrostatic contributions compliant with the ideas of the IEM web service, and is thus well suited to serve as its default method for the calculation of electrostatics interactions. It should be noted for completeness that, when the GB model is used, the IEs provided by the IEM web service constitute effective energies affected by the implicit solvent; more specifically, while the IEs remain pairwise separable, they are affected by the calculated Born radii, which depend not only on the configuration of the residue pair, but also on the global geometry of the biomolecular 3D structure.

Implementation of the GB model has required the extension of the computational pipeline to allow for the Born radii calculation. The calculation is performed after the atom types and FF parameters had been assigned, since the radii are atom type-specific. The required parameters have been adopted from the GROMACS simulation suite (18) for all supported FFs.

Information content (conservation) calculation

After selecting a biomolecular structure, the PDB file is processed using the Bio.PDB module (19) from the Biopython package (20) and sequences of its individual chains are read. The polypeptide chains are then identified, and a MSA of similar sequences is constructed for each using the *phmmer* tool from the HMMER software package (http://hmmer.org/). The *phmmer* search uses default parameters and UniProtKB/Swiss-Prot (21) as the target database. The target database will continue to be updated and its current version can be found in the IEM web service manual. For each MSA, the respective weights of its constituting sequences are calculated using the Gerstein/Sonnhammer/Chothia algorithm (22) as implemented in the *esl-weight* miniapp included with the HMMER software. Finally, the per-residue IC is calculated using the *esl-alistat* miniapp, taking the individual sequence weights into account.

Unlike ConSurf (3), the conservation calculation pipeline does not attempt to reconstruct the phylogenetic tree or model the sequence evolution, relying instead on the mentioned sequence-weighting algorithm to reduce the effects of uneven phylogenetic representation bias. Together with performing only pairwise alignments of the identified similar sequences against the query, this leads to a massive speed-up of the calculations, with most queries having the IC values calculated in a matter of seconds.

The IC values range between 0.0 bits (no conservation) and approx. 4.32 [$\log_2(20)$] bits (complete conservation). Please be aware that large IC values are also occasionally exhibited by the chains' terminal residues, as these can remain unmatched in the MSAs. Identification of such artifacts may be enabled in the future with the utilization of the respective MSA per-column gap frequency values.

The conservation calculation pipeline is implemented as a Python script and is publicly available at https://github.com/davidjakubec/INTAA-conservation under a permissive license. The pipeline is invoked automatically for each new task started by the IEM web service in parallel with the IE calculations and, due to the speed of the HMMER software and the size of the target database, constitutes at worst only a negligible slowdown of the overall computations.

New visualization options

In the current version of the application, the IC values for individual amino acid residues have been added numerically to the results table, and are also being visualized in the first column of the table using green bars (Figure 1). The current version also adds a simple sequence visualization window (*Sequence* tab) in which individual residues are colored according to their respective total IEs and IC values. The sequence visualization window is interactive and allows details about individual residues to be displayed by hovering the mouse cursor over them (Figure 1).

The integrated molecular structure viewer allows the visualized 3D structure to be annotated with an IC-based mark-up by selecting the option *conservation* from the *Selection* drop-down menu (Figure 1).

In the current version of the application, the pairwise IE as well as residue distance matrices can be visualized in-browser using the *Residue Matrix* tab by selecting the options *energy matrix* or *distance matrix*, respectively, in the *Selection* drop-down menu (Figures 1 and 2). The distance between a pair of residues can be calculated as either the closest distance between any pair of their atoms, or as the distance between the centers of mass (COMs) of the considered moieties; these options can be toggled using the *Configure* dialog in the new *Action* drop-down menu (Figure 1). The residue distance calculations reflect the subgroup selection in the *Interaction* drop-down menu; for example, when the IEs are calculated only for the interactions among amino acid backbone atoms, only these are considered in the residue distance calculation. Finally, the in-browser matrix viewer can also visualize a combination of pairwise IEs and residue distances by selecting the option *combined matrix*. In this mode, residue pairs with mutual distance below a threshold are colored according to the pairwise IE, while pairs further apart are shown masked in greyscale. The motivation behind this mode is that the IEs for the interactions among distant residues are usually close to zero and are thus hidden due to being uninformative. The matrix viewer is interactive and allows information about the pairwise IEs and residue distances to be shown by hovering the mouse cursor over the individual residue pairs (Figure 2).

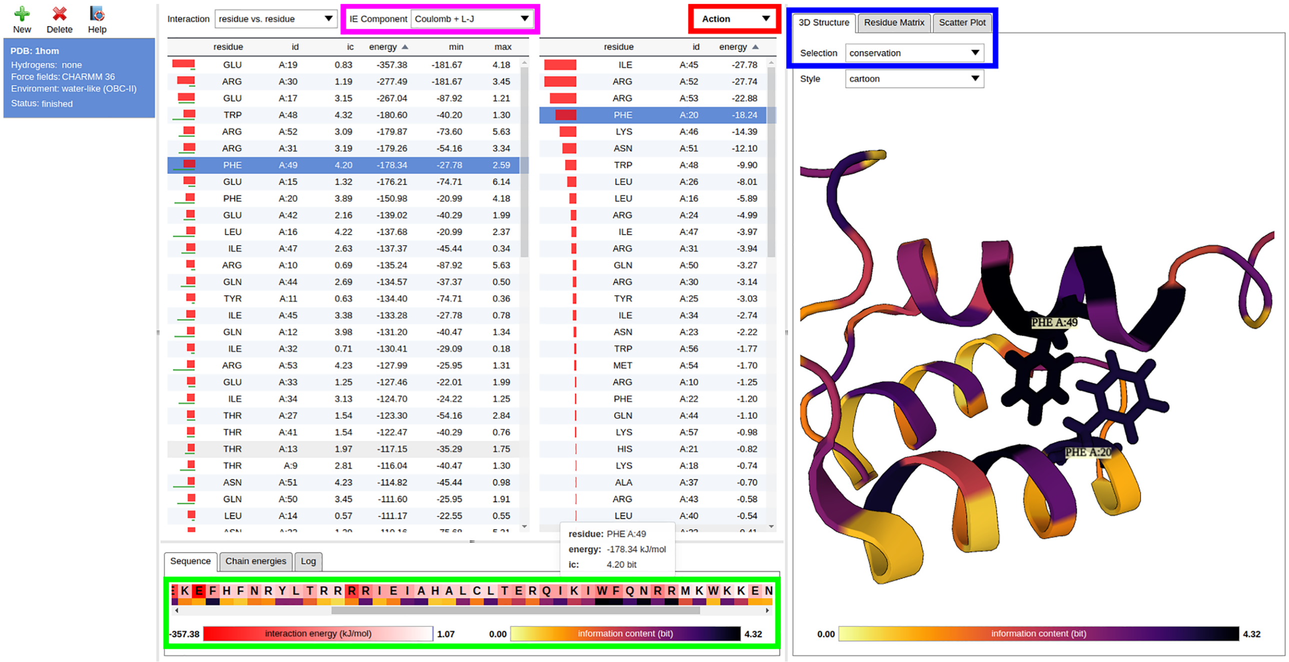
The in-browser matrix viewer can also visualize the pairwise IE and residue distance matrices for only selected parts (e.g. chains) of the investigated structure. These can be specified individually for the matrix' rows and columns using the *Residues* drop-down menus. The size of the matrix elements can be configured using the *Size* drop-down menu (Figure 2).

The color scale limits and selected thresholds for the respective viewers utilized in the *3D Structure* and *Residue Matrix* tabs can be newly configured using the new *Action* drop-down menu (Figure 1).

Finally, a new *Scatter Plot* tab has been added, providing a simple way to interrelate the total IEs and IC values for the selected residues in the form of an interactive two-dimensional (2D) plot (Figure 3).

Other changes and improvements

A new *IE Component* drop-down menu has been added, allowing the user to choose between the display of the Coulomb, Lennard-Jones, or complete (i.e. Coulomb + Lennard-Jones) IEs (Figure 1). Next, the front page of the INTAA web server has received a major graphical overhaul which improves its clarity and highlights the most important information. In addition, every element in the IEM web service user interface has been extended with an informative tooltip which provides instantaneous access to its detailed description. The tooltips are accessible by simply hovering the mouse cursor over the respective elements.



PDB: 1hom
Hydrogens: none
Force fields: CHARMM 36
Enviroment: water-like (OBC-II)
Status: finished

Interaction: residue vs. residue
IE Component: Coulomb + L-J

residue	id	ic	energy ▲	min	max
GLU	A:19	0.83	-357.38	-181.67	4.18
ARG	A:30	1.19	-277.49	-181.67	3.45
GLU	A:17	3.15	-267.04	-87.92	1.21
TRP	A:48	4.32	-180.60	-40.20	1.30
ARG	A:52	3.09	-179.87	-73.60	5.63
ARG	A:31	3.19	-179.26	-54.16	3.34
PHE	A:49	4.20	-178.34	-27.78	2.59
GLU	A:15	1.32	-176.21	-74.71	6.14
PHE	A:20	3.89	-150.98	-20.99	4.18
GLU	A:42	2.16	-139.02	-40.29	1.99
LEU	A:16	4.22	-137.68	-20.99	2.37
ILE	A:47	2.63	-137.37	-45.44	0.34
ARG	A:10	0.69	-135.24	-87.92	5.63
GLN	A:44	2.69	-134.57	-37.37	0.50
TYR	A:11	0.63	-134.40	-74.71	0.36
ILE	A:45	3.38	-133.28	-27.78	0.78
GLN	A:12	3.98	-131.20	-40.47	1.34
ILE	A:32	0.71	-130.41	-29.09	0.18
ARG	A:53	4.23	-127.99	-25.95	1.31
GLU	A:33	1.25	-127.46	-22.01	1.99
ILE	A:34	3.13	-124.70	-24.22	1.25
THR	A:27	1.54	-123.30	-54.16	2.84
THR	A:41	1.54	-122.47	-40.29	0.76
THR	A:13	1.97	-117.15	-35.29	1.75
THR	A:9	2.81	-116.04	-40.47	1.30
ASN	A:51	4.23	-114.82	-45.44	0.98
GLN	A:50	3.45	-111.60	-25.95	1.91
LEU	A:14	0.57	-111.17	-22.55	0.55

residue	id	energy ▲
ILE	A:45	-27.78
ARG	A:52	-27.74
ARG	A:53	-22.88
PHE	A:20	-18.24
LYS	A:46	-14.39
ASN	A:51	-12.10
TRP	A:48	-9.90
LEU	A:26	-8.01
LEU	A:16	-5.89
ARG	A:24	-4.99
ILE	A:47	-3.97
ARG	A:31	-3.94
GLN	A:50	-3.27
ARG	A:30	-3.14
TYR	A:25	-3.03
ILE	A:34	-2.74
ASN	A:23	-2.22
TRP	A:56	-1.77
MET	A:54	-1.70
ARG	A:10	-1.25
PHE	A:22	-1.20
GLN	A:44	-1.10
LYS	A:57	-0.98
HIS	A:21	-0.82
LYS	A:18	-0.74
ALA	A:37	-0.70
ARG	A:43	-0.58
LEU	A:40	-0.54

residue: PHE A:49
energy: -178.34 kJ/mol
ic: 4.20 bit

E K E F H F N R Y L T R R R R I E I A H A L C L T E R Q I K I W F Q N R R M K W K K E N

Selection: conservation
Style: cartoon

Figure 1. The 3D structure of the *Antennapedia* homeodomain from *Drosophila* (PDB ID 1HOM) analyzed using the IEM web service. As the hydrogen atoms were already present in the structure, none needed to be added. The IEs were calculated using the CHARMM36 FF (25) and the new OBC-II continuum solvation model. The central section of the UI features two tables of IE values. In the table on the left, the amino acid residues have been sorted according to their total IEs, with the ones displaying the most energetically favorable interactions at the top [the residues can be sorted according to any of the displayed features as described in the previous publication (5)]. The IC values calculated for each residue using the new conservation calculation pipeline have been added to this table. Once a residue is selected (in this example, Phe49), the table on the right side shows the pairwise IEs for its interactions with the other residues in the structure. These IEs can be sorted as well and the most favorable pairwise interactions can thus be identified for each residue. In this example, Phe20 is selected in the table on the right. It can be seen that the pairwise noncovalent interaction between the two selected residues is among the most energetically favorable. The integrated molecular structure viewer in the right section of the UI displays the investigated structure in cartoon representation; any residues selected within the central tables are highlighted in stick representation. The most notable additions to the web service and UI updates are highlighted in bold rectangles. The top magenta rectangle highlights the new *IE Component* drop-down menu. Here, the user can choose between the display of the Coulomb, Lennard-Jones, or complete (Coulomb + Lennard-Jones) IEs in the application. The bottom green rectangle highlights a window relating the primary sequence, the one-dimensional (1D) IE profile, and the 1D conservation profile for the residues in the structure. The window is interactive and enables the IE and IC values for individual residues to be displayed numerically. The top right blue rectangle highlights the selection boxes for the new visualization options. The *Selection* drop-down menu in the *3D Structure* tab allows the user to switch between total IE-, pairwise IE-, or conservation-based mark-ups for the investigated structure. In this example, *conservation* is selected, and the per-residue IC values are projected onto the structure. It can be seen that the two selected Phe residues are involved in the hydrophobic core of the homeodomain and are both highly conserved, in agreement with the known evolutionary properties of the domain family (24). Switching to the *Residue Matrix* tab allows the user to visualize and interact with the IE and distance matrices; switching to the *Scatter Plot* tab allows the user to interrelate the total IEs and IC values for the selected residues using an interactive 2D plot. These are described in greater detail in Figures 2 and 3, respectively. Finally, the top red rectangle highlights the new *Action* drop-down menu. This menu provides access to advanced IEM export options, configuration options, as well as a button for providing feedback to the developers.

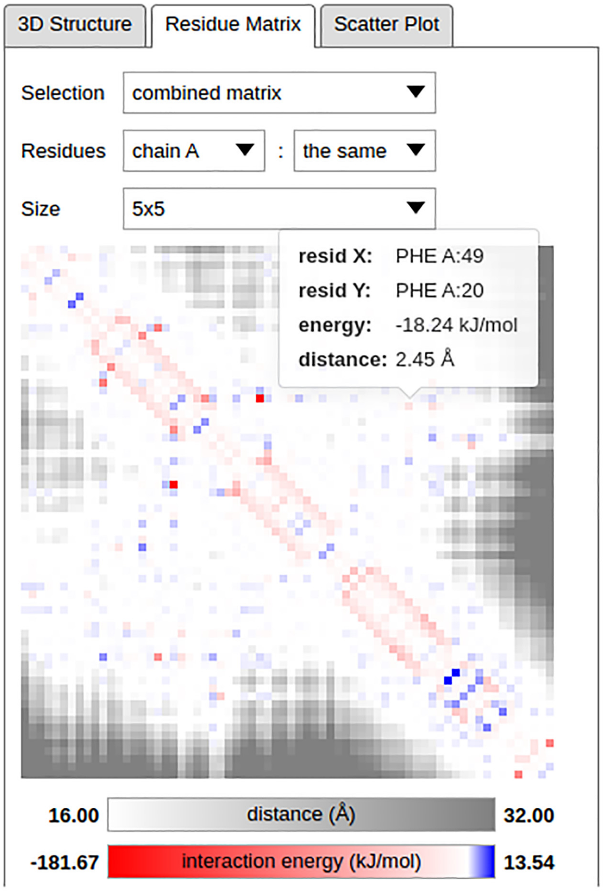
Furthermore, the user's choices for the drop-down menu selections are newly remembered across sessions. Finally, the on-line manual page of the IEM web service has received a major update to its content and structure and now also includes illustrative examples of the application's usage.

RESULTS

We demonstrate the new functionality of the INTAA web server and the types of analyses it enables using an example task featuring the 3D structure of the *Antennapedia* homeodomain from *Drosophila* deposited in the PDB under the ID 1HOM (23). The structure of this DNA-binding protein domain was determined using nuclear magnetic resonance spectroscopy in solution and constitutes a suitable demonstrative example due to its small size and well-known structure–function relationship and conservation profile (24). The analysis of this structure enabled by the INTAA web server is shown in Figure 1. This Figure reviews selected elements of the IEM web service and illustrates how novel information about the relationship between energetic and evolutionary roles of specific residues can be obtained with the utilization of the new web server features. Figure 2 follows this analysis with a detailed view of the new, interactive IE and distance matrix visualization for the investigated structure. Finally, Figure 3 shows the total IE–IC value relations for all residues in the investigated structure in the form of a new, interactive 2D scatter plot visualization.

DISCUSSION

As described above, the GB model introduced into the IEM web service in this update improves the quality and accuracy of the IE calculations. Nevertheless, the significance of the IE calculations can still depend on the quality of



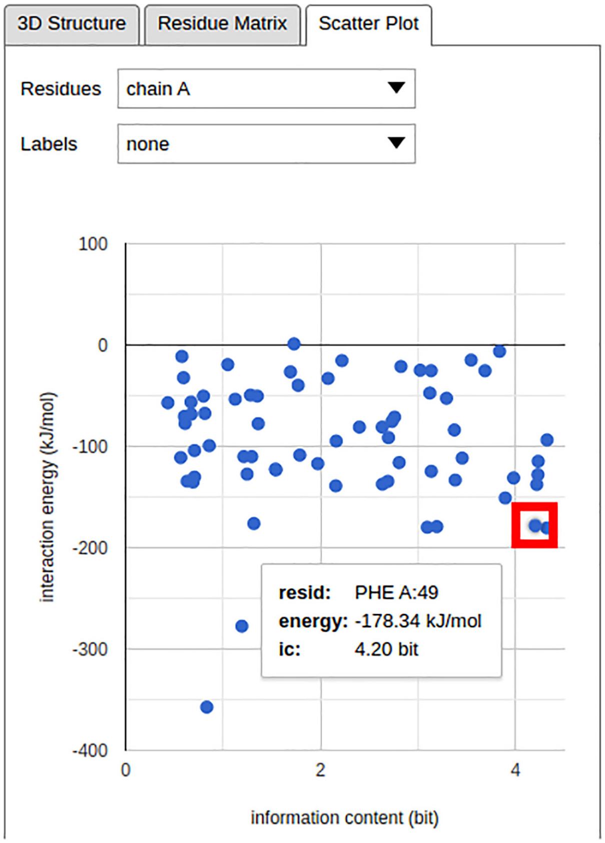
Selection: combined matrix

Residues: chain A : the same

Size: 5x5

resid X: PHE A:49
resid Y: PHE A:20
energy: -18.24 kJ/mol
distance: 2.45 Å

Figure 2. Detailed view of the IE and distance matrices corresponding to the structure investigated in Figure 1. The *combined matrix* has been chosen for visualization in the *Selection* drop-down menu, meaning that IEs are shown in the matrix viewer for residue pairs with any pair of atoms at or below a distance threshold and distances are shown for residue pairs further apart. An alternative, COM-based distance definition can be toggled using the *Configure* dialog in the new *Action* drop-down menu (Figure 1); the thresholds can be configured using this menu as well. The investigated structure contains only a single chain and the *Residues* drop-down menu selections thus lead to a symmetric matrix being displayed. Finally, the *Size* drop-down menu can be used to adjust the dimensions of the matrix elements. In this example, the Phe49–Phe20 pair investigated in Figure 1 has been highlighted in the matrix viewer, and the IE for its pairwise interaction and the closest distance between the residues' atoms are shown.



Residues: chain A

Labels: none

resid: PHE A:49
energy: -178.34 kJ/mol
ic: 4.20 bit

Figure 3. The total IE–IC value relations for the residues in the structure investigated in Figure 1. The single chain in the investigated structure has been selected using the *Residues* drop-down menu; the *Labels* drop-down menu can be used to add textual identifiers to individual points in the plot. In this example, the position of Phe49 in the scatter plot has been highlighted in a red rectangle, and its total IE and IC value are shown.

the user's input. Two major sources of input-related errors are known at this time. First, errors in the macromolecular structures, such as missing nonhydrogen atoms and unrealistically close atomic contacts, can result in missing or unphysical elements in the IEM. These propagate into other quantities, such as total IEs. Second, unrealistic atomic overlaps and other errors can also result from the use of tools utilized to add the missing hydrogen atoms which are necessary for the IE calculations. Both of these issues could be remedied by additional processing of the investigated structures, including modeling of its missing parts and subsequent geometry optimization. However, these procedures are difficult to implement and to perform automatically and would likely become major sources of errors or unstable behavior themselves. In addition, they would significantly increase the computational demands of the application and introduce an undesirable delay between the task submission and presentation of the results. Therefore, we decided to omit these steps and encourage the users to carefully inspect the biomolecular structures for potential errors before utilizing the IEM web service.

In this update, we have described major additions to the INTAA web server and, in particular, to the IEM web service. The utilization of the GB model makes the IE calculations in aqueous media more physically accurate. The addition of the conservation score calculations makes it, for the first time, possible to study the residues' physical and evolutionary roles within the biomolecular structures using a single web service. The INTAA web server remains aimed for use by bioinformaticians as well as biochemists and, as illustrated in this work, the new features enable it to be utilized for a wide range of tasks in biomolecular structure analysis and engineering which benefit from interrelating the physical and evolutionary views of the structures.

DATA AVAILABILITY

The INTAA web server is publicly available at https://bioinfo.uochb.cas.cz/INTAA/. The source code for carrying out the IC calculation tasks is available at https://github.com/davidjakubec/INTAA-conservation.

FUNDING

ELIXIR CZ research infrastructure project: Ministry of Education, Youth and Sports [LM2018131] including access to computing and storage facilities; European Regional Development Fund/European social fund project 'Chemical biology for drugging undruggable targets (ChemBioDrug),' [CZ.02.1.01/0.0/0.0/16_019/0000729]; Czech Science Foundation [GA19-03488S]. Funding for open access charge: Ministry of Education, Youth and Sports [LM2018131].

Conflict of interest statement. None declared.

REFERENCES

1. Bendová-Biedermannová,L., Hobza,P. and Vondrášek,J. (2008) Identifying stabilizing key residues in proteins using interresidue interaction energy matrix. *Proteins*, **72**, 402–413.
2. Vondrášek,J., Bendová,L., Klusák,V. and Hobza,P. (2005) Unexpectedly strong energy stabilization inside the hydrophobic core of small protein rubredoxin mediated by aromatic residues: correlated ab initio quantum chemical calculations. *J. Am. Chem. Soc.*, **127**, 2615–2619.
3. Celniker,G., Nimrod,G., Ashkenazy,H., Glaser,F., Martz,E., Mayrose,I., Pupko,T. and Ben-Tal,N. (2013) ConSurf: using evolutionary data to raise testable hypotheses about protein function. *Isr. J. Chem.*, **53**, 199–206.
4. Jakubec,D., Vondrášek,J. and Finn,R.D. (2019) 3DPatch: fast 3D structure visualization with residue conservation. *Bioinformatics*, **35**, 332–334.
5. Galgonek,J., Vymetal,J., Jakubec,D. and Vondrasek,J. (2017) Amino Acid Interaction (INTAA) web server. *Nucleic. Acids. Res.*, **45**, W388–W392.
6. Berman,H., Henrick,K. and Nakamura,H. (2003) Announcing the worldwide protein data bank. *Nat. Struct. Mol. Biol.*, **10**, 980–980.
7. Mirzaie,M. (2019) Discrimination power of knowledge-based potential dictated by the dominant energies in native protein structures. *Amino Acids*, **51**, 1029–1038.
8. Barik,S. (2020) The uniqueness of tryptophan in biology: properties, metabolism, interactions and localization in proteins. *Int. J. Mol. Sci.*, **21**, 8776.
9. Chakrabarty,B., Naganathan,V., Garg,K., Agarwal,Y. and Parekh,N. (2019) NAPS update: network analysis of molecular dynamics data and protein-nucleic acid complexes. *Nucleic Acids Res.*, **47**, W462–W470.
10. Musyoka,T.M., Njuguna,J.N. and Bishop,Ö.T. (2019) Comparing sequence and structure of falcipains and human homologs at prodomain and catalytic active site for malarial peptide based inhibitor design. *Malar. J.*, **18**, 159.
11. Chakravarty,D., Bihani,S.C., Banerjee,M. and Ballal,A. (2019) Novel molecular insights into the anti-oxidative stress response and structure–function of a salt-inducible cyanobacterial Mn-catalase. *Plant Cell Environ.*, **42**, 2508–2521.
12. Alhamoudi,K.M., Bhat,J., Nashabat,M., Alharbi,M., Alyafee,Y., Asiri,A., Umair,M. and Alfadhel,M. (2020) A missense mutation in the UGDH gene is associated with developmental delay and axial hypotonia. *Front. Pediatr.*, **8**, 71.
13. Grochowski,P. and Trylska,J. (2008) Continuum molecular electrostatics, salt effects, and counterion binding—a review of the Poisson–Boltzmann theory and its modifications. *Biopolymers*, **89**, 93–113.
14. Riley,K.E., Vondrášek,J. and Hobza,P. (2007) Performance of the DFT-D method, paired with the PCM implicit solvation model, for the computation of interaction energies of solvated complexes of biological interest. *Phys. Chem. Chem. Phys.*, **9**, 5555–5560.
15. Onufriev,A. and Case,D.A. (2019) Generalized born implicit solvent models for biomolecules. *Annu. Rev. Biophys.*, **48**, 275–296.
16. Onufriev,A., Case,D.A. and Bashford,D. (2002) Effective born radii in the generalized Born approximation: the importance of being perfect. *J. Comput. Chem.*, **23**, 1297–1304.
17. Onufriev,A., Bashford,D. and Case,D.A. (2004) Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins Struct. Funct. Genet.*, **55**, 383–394.
18. Abraham,M.J., Murtola,T., Schulz,R., Páll,S., Smith,J.C., Hess,B. and Lindahl,E. (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, **1–2**, 19–25.
19. Hamelryck,T. and Manderick,B. (2003) PDB file parser and structure class implemented in Python. *Bioinformatics*, **19**, 2308–2310.
20. Cock,P.J.A., Antao,T., Chang,J.T., Chapman,B.A., Cox,C.J., Dalke,A., Friedberg,I., Hamelryck,T., Kauff,F., Wilczynski,B. and de Hoon,M.J.L. (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, **25**, 1422–1423.
21. The UniProt Consortium (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic. Acids. Res.*, **47**, D506–D515.
22. Gerstein,M., Sonnhammer,E.L.L. and Chothia,C. (1994) Volume changes in protein evolution. *J. Mol. Biol.*, **236**, 1067–1078.
23. Billeter,M., Qian,Y.-Q., Otting,G., Müller,M., Gehring,W.J. and Wüthrich,K. (1990) Determination of the three-dimensional structure of the *Antennapedia* homeodomain from *Drosophila* in solution by ^{1}H nuclear magnetic resonance spectroscopy. *J. Mol. Biol.*, **214**, 183–197.
24. Bürglin,T.R. and Affolter,M. (2016) Homeodomain proteins: an update. *Chromosoma*, **125**, 497–521.
25. Huang,J., Rauscher,S., Nawrocki,G., Ran,T., Feig,M., de Groot,B.L., Grubmüller,H. and MacKerell,A.D. Jr (2017) CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods*, **14**, 71–73.