Assessment of Semiempirical Enthalpy of Formation in Solution as an Effective Energy Function to Discriminate Native-like Structures in Protein Decoy Sets

Gabriel Aires Urquiza-Carvalho, Wallace Duarte Fragoso, and Gerd Bruno Rocha*

In this work, we tested the PM6, PM6-DH+, PM6-D3, and PM7 enthalpies of formation in aqueous solution as scoring functions across 33 decoy sets to discriminate native structures or good models in a decoy set. In each set these semiempirical quantum chemistry methods were compared according to enthalpic and geometric criteria. Enthalpically, we compared the methods according to how much lower was the enthalpy of each native, when compared with the mean enthalpy of its set. Geometrically, we compared the methods according to the fraction of native contacts (Q), which is a measure of geometric closeness between an arbitrary structure and the native. For each set and method, the Q of the best decoy was compared with the Qv, which is the Q of the decoy closest to the native in the set. It was shown that the PM7 method is able to assign larger energy differences between the native structure and the decoys in a set, arguably because of a better description of dispersion interactions, however PM6-DH+ was slightly better than the rest at selecting geometrically good models in the absence of a native structure in the set. © 2016 The Authors. Journal of Computational Chemistry Published by Wiley Periodicals, Inc.

DOI: 10.1002/jcc.24415

Introduction

Protein tertiary structure prediction is an obstacle in molecular biology that has challenged researchers in chemistry, biology, and physics for over half a century.11–13 Traditional approaches have attempted to find suitable methods for conformational sampling and effective energy functions that can correctly identify native-like conformations.

Even though the large size of the search space and the complexity of the energy landscape make extensive sampling and global energy optimization essentially impossible to use, physics-based effective energies for most biomolecular systems, these energies represent a very important tool for scoring and refining protein models produced by heuristic methods, such as Rosetta,4,5 TASSER,6–8 3D-SHOOTGUN,9 and so forth.

The methods above were shown to produce sets of models that contain relatively accurate native-like conformations among non-native decoy conformations but they are usually unable to reliably discriminate the native conformations among a set of decoys. These decoy sets are created using a variety of sampling algorithms.8,10–13

Because protein structure prediction is usually based on the Anfinsen thermodynamic hypothesis,14 a crucial property for an effective energy function is its ability to recognize when a conformation is at a free energy minimum. This means that a good test for an effective energy function is to score a large number of closely related structures to recognize native-like conformations, a purpose for which the decoy sets produced by the methods cited above are widely used.15–18

The correct assignment of an effective energy to a conformation requires a good physical description of the stability of the chain, and nonbonded interactions play a decisive role in the conformational stability of proteins, particularly hydrogen bonds22 and hydrophobic interactions. Despite the fact that a number of force fields have been parametrized to describe these interactions with a good margin of accuracy, the correct description of these types of interactions can only be done through the recovery of a large portion of electron correlation using large basis sets and high-level quantum chemistry methods.

To avoid such costly computational solutions, ad hoc corrections in the energetics of dispersion interactions were proposed by Grimme19,20 for DFT methods, followed closely by McNamara and Hillier21 and more recently by Anikin22 for semiempirical quantum chemical methods. Similar corrections for hydrogen bonds also followed, using specific methodologies.23–25 The accuracy of semiempirical methods corrected for hydrogen bonds and dispersion interactions has been reported to be comparable to DFT corrected for dispersion and up to three times faster.26

The corrections to noncovalent interactions in semiempirical methods has been thoroughly tested in biological systems.27–29 In particular, the PM6 method30 and its family of derived methods have been widely used to model biological and other complex systems on several occasions.28,31–34

The PM6 method has also been used to model the geometries of proteins with a good degree of success.35 Recently, PM6-DH+ has also been highlighted by Yilmazer et al. as a commercial purpose. Attribution, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2016 The Authors. Journal of Computational Chemistry Published by Wiley Periodicals, Inc.
fast and accurate alternative to a full ab initio computation of the interactions between proteins and their ligands.\textsuperscript{[38]}

Maia and coworkers used the PM6-DH method\textsuperscript{[24]} to assess the performance and accuracy of MOPAC2012 GPU-implementation\textsuperscript{[37]} using the software to make implicit solvent calculations for the enthalpies of formation of three decoy sets taken from the I-TASSER decoy set II\textsuperscript{[10]} assembly. In these calculations, the native structures in each set had a markedly different enthalpy of formation in relation to the rest of the conformations in the set, indicating the possibility of using the semiempirical enthalpies of formation as an effective energy function to distinguish native-like structures from the rest of the decoy set.

This marked difference in the enthalpy of the native structure when compared to the rest led us to the hypothesis that perhaps the surface of the enthalphy of formation in solution around the native structure might be unusually steep.

We decided to test whether the enthalpies of formation in solution, calculated using the PM6/PM7 family of semiempirical Hamiltonians, can be used to find native-like conformations in a set of decoys. We also want to investigate the effect that corrections to noncovalent interactions to the final enthalpy of formation have in that scoring capacity, if any.

Using enthalpies of formation as basis for scoring decoy sets, Wollacott and Merz\textsuperscript{[38]} compared the results of $\Delta G$ estimations using the molecular mechanics AMBER, CHARMM, and OPLS force fields as well as the semiempirical PM3 and AM1 Hamiltonians. Some of the decoy sets used in their work are also present in our assembly.

In their work, Wollacott and Merz applied the semiempirical enthalpies of formation from the AM1 and PM3 Hamiltonians to the DivCon score, shown in eq. (1), along with the polarization energy, obtained from solving the finite difference Poisson equation and the attractive Lennard–Jones term taken from the AMBER parm94 force field.

$$\Delta G = a\Delta H + b\Delta G_{sol} + c\Sigma LJ$$  \hspace{1cm} (1)

In this equation, the term $\Delta G$ is the total free energy of the system. $\Delta G_{sol}$ is the free energy of solvation, $\Delta H_{LJ}$ is the enthalpy of formation as calculated by a Hamiltonian (either AM1 or PM3 in Wollacott’s and Merz’s work) and the term $\Sigma LJ$ represents the attractive Lennard–Jones term. Since this estimation mixes terms calculated at different levels of theory, the weights $a$, $b$, and $c$ are introduced and assigned in a way to maximize the Z-Score (a measure of distance from the structure’s score to the mean score of the population) difference between the native structure and the decoys.

Prior to their scoring, the structures in the sets were minimized using the AMBER force field. The metrics used in their assessment included the Z-Score gap and overall ranking. In their work, they showed that the DivCon score using the AM1 and PM3 semiempirical Hamiltonians is consistently better at scoring the native structure than the AMBER parm94 molecular mechanics force field. Furthermore, the PM3 Hamiltonian was shown to be slightly better than the AM1 Hamiltonian.

In their analysis, Wollacot and Merz compared how good the final free energy estimations, as calculated by eq. (1) (where $\Delta H_{LJ}$ is either PM3 or AM1 Hamiltonian), were at ranking the native structures. They also showed the results for the enthalpies of formation in gas phase, which were much worse. However, Maia and his collaborators\textsuperscript{[37]} showed that the enthalpies of formation in solution on their own are able to detect the native structure in sets of decoys.

Merz and Wollacot’s work also points to the fact that in their sets of decoys, although the native structure usually had a good score in terms of heat of formation, they all scored poorly in respect to solvation, when compared to the decoys. This trend is also seen in the best scoring decoys. All native structures in their work also had a favorable Lennard–Jones term, suggesting that tight packing might compensate for the reduction in the free energy of solvation. Overall, both AMBER and DivScore demonstrate poor discrimination of the decoys by score as a function of RMSD. However, the work called our attention to the fact that since the protein decoys have been created computationally using various force fields, scoring the decoy structures relative to the unmodified crystal structure introduces bias. The decoys may have geometries that are better suited to the force field, improving their score relative to the native structure.

A few years later, Scheraga et al. released a paper in which they studied a different way of using decoys to test a scoring function. Instead of minimizing the decoys, they relaxed the structures with short Monte Carlo runs.\textsuperscript{[39]}

They tested three energy functions, all of which were a combination of an all-atom force field and a solvation term: ECEPP05 combined with a solvent-accessible surface area model with the parameters optimized with a set of protein decoys (ECEPP05/SA); ECEPP/3 combined with a different solvent-accessible surface area model (ECEPP3/OONS) and ECEPP05 combined with an implicit solvent model based on a solution of the Poisson equation with an optimized Fast Adaptive Multigrid Boundary Element (FAMBEpH) method (ECEPP05/FAMBEpH).

Scheraga and collaborators\textsuperscript{[39]} then compared these three functions for their ability to rank the native structures on 45 of all the proteins in the Rosetta decoy sets after a Monte Carlo with Minimization (MCM) run and a local energy minimization. They compared this with the scoring method using only local minimization. To assess the reliability of the methods they compared the scores of the 5 and 10 structures with the lowest $C^*$ RMSD from the native structure to assess the method’s ability to adequately rank near-native structures.

In their analysis, they showed that the ECEPP05/SA combination was particularly good at ranking the native structures on 45 of all the proteins in the Rosetta decoy sets after a Monte Carlo with Minimization (MCM) run and a local energy minimization. Nevertheless, the ECEPP05/FAMBEpH combination was the overall best combination, trailing behind the ECEPP05/SA combination in terms of practicality simply because its solvation term is more complicated to compute. In all cases, the discriminative power increased when some form of simulation was used to relax the structures found in the decoy set prior to scoring.
Lazaridis and Karplus also studied ways of estimating effective energy functions for protein conformational studies. In their work, they compared the ability of a CHARMM potential to discriminate between the native structure and its decoys with and without a Gaussian estimation for the Gibbs free energy variation due to the solvation, ΔG_solv.

The publication of the PM7 method and its new parameterization scheme for dispersion interactions as well as the work of Merz et al. and that of Lazaridis and Karplus motivated our revisitation of the decoy sets, this time using the PM7 Hamiltonian and the COSMO solvation model. The use of the PM3 method in Wollacott and Merz’s work, a somewhat old method in the PMx family of semiempirical Hamiltonians, and the absence of dispersion corrections in their work motivated us to build on their steps, assessing the ability of the enthalpy of formation of the PM7 method in solution to discriminate native-like structures from a set of decoy models.

The idea was to do a different approach on the notion of Merz and Wollacot’s work instead of exploring the mixing of molecular mechanics and semiempirical methods as a way to estimate free energy, we decided to use only the enthalpies of formation in aqueous solution calculated by semiempirical quantum chemistry methods.

The reason is that not only are the semiempirical enthalpies of formation fully quantum mechanical descriptions of the system, they are also quite fast to compute. Furthermore, their versions corrected for noncovalent interactions have been shown to be comparably accurate to DFT ab initio calculations.

Because Maia and his collaborators used the PM6-DH+ method in the original work on GPU implementation and the PM6-D3 correction scheme is an improvement over the DH+, we decided to apply both the PM6-DH+ and PM6-D3 along with PM7 methods to the entire ensemble of decoy sets, instead of only the three sets presented in that paper.

To compare the PM7, PM7, PM6-DH+ and PM6-D3 enthalpies of formation in solution as effective energy functions to find native structures in the decoy sets, we needed a kind of “control” energy function that lacked any corrections whatsoever. This energy was the PM6 method, which itself has no corrections for dispersion or hydrogen bond interactions.

Because the trend we observed was in the enthalpies of formation, we decided not to estimate the free energy but rather to determine whether the enthalpy of formation in solution could be used to separate the native-like structure from those which are not native-like and furthermore to determine how the dispersion corrections affected accuracy.

Methodology

The decoy set was downloaded from the Zhang Lab site, decoy set type II, generated by the I-TASSER Algorithm. This assembly contains 56 sets but not all were used. For some sets, the native structures was missing entire side chains or atoms in the amino-terminus or had other structural imperfections and were therefore abandoned.

For the remaining 33 sets (Table 1), the UCSF Chimera software was used to add hydrogen atoms to the native structures. The protonation states of Histidines were determined by Chimera, which took into consideration steric limitations and any possible formation of hydrogen bonds. Subsequently, valences and disulfide bridges were verified using the information in the PDB file headers using the SCIGRESS software.

Hydrogen-only optimizations were carried out using MOPAC separately for each method. Because only the hydrogen atoms were optimized, a convergence threshold for the gradient norm was set to 1.0 kcal mol⁻¹ Å⁻¹. For multichain proteins, only chain A was used.

The molecules were grouped based on their single-point heats of formation in aqueous solution. In total, we carried out 61280 semiempirical quantum chemistry calculations across 33 decoy sets using the MOPAC software and the COSMO implicit solvent field with a relative permittivity of 78.4 and an effective solvent molecule radius of 1.3 Å.

The calculations were run using the linear-scaling technique MOZYME. For the hydrogen optimization in the native structures, the MOZYME cutoff radius was 10 Å, and for the single-point calculations, it was set to 9 Å. After the calculations, decoys that showed large positive enthalpies of formation were excluded from further analysis because it was verified that the SCF (Self-Consistent Field) did not converge for such cases. Table 1 summarizes the overall characteristics of the sets considered in this work.

It has been known for a long time that the native ensemble of a protein is composed of structures that are close together on a range of about 2 Å. It is also been known for a long time that the entropy of the native state can be approximated by contributions from the small number of native conformations in relation to the vast amount of unfolded conformations.

This means that the native state must be located at a very deep minimum in the energy hypersurface and thus there is a significant energy gap between a native and an average non-native structure.

Because of this, one can expect a good score function to have this property. To assess the efficacy of semiempirical methods as scoring functions, we tested whether they satisfy these energy gap criteria.

We assumed that a random ensemble of conformations have a Gaussian distribution of enthalpies. Therefore, they belong to a population of structures with similar geometric and energetic properties. This is not an unreasonable assumption, since the sampling algorithm usually explores a small area of the phase space. However, certain conformations have extremely different enthalpy values. For these cases, we assumed that these specific conformations are outliers that differ from the population.

High-enthalpy conformations occur systematically and have no physical meaning. However, the presence of these outliers can strongly impact the ensemble average because they often have extreme values. Thus, the first step is to delete all high-enthalpy conformations. To do this, a limit of two standard deviations above the mean was established and the conformations with a heat of formation that exceeds this limit are
removed from the ensemble. Then, the mean is recalculated and a new limit is set, and the process is repeated until there are no more high-enthalpy conformations.

After this iterative process concludes due to lack of outliers in the remaining set, a new three standard deviation limit is set below the mean. Conformations with enthalpy values lower than this limit are considered native-like. Notably, the lower limit (3s) is wider than the upper limit (2s), because while the upper limit excludes conformations without physical interpretation, the lower limit indicate native-like conformations and is more stringent.

The complete algorithm is shown in Figure 1. This assessment allows us to verify whether a method satisfies the energy gap property that an energy function must have to correctly describe the energy hypersurface. The methods that managed to assign such markedly lower scores to the native structure were considered better.

The results and discussion section outlines the methods and decoy sets used in the study. The decoy sets are characterized in Table 1. The methods are then tested in two ways: (1) their capacity to detect the native structure in the set and (2) the overall quality of the models selected in the absence of the native structure.

To assess the method’s overall ability to rank the native structures correctly we used the Z-score shown in eq. (2), which was also used in Wallacott and Merz’s work. In this equation, $H$ is the mean enthalpy of the population, $s$ is the standard deviation, and $H$ is the enthalpy of a particular conformation in the set.

$$Z(H) = \frac{H - \bar{H}}{s}$$

Here we show $Z(H_{native})$ for all sets and methods to see how far the enthalpy of the native structure is from the mean enthalpy. This is a very general quality measure for a score function because as the native structure gets further from the mean structure, it becomes increasingly more likely that it is the structure with lowest enthalpy. These data are summarized in Table 2.

As shown in Table 2, the PM7 method had a better overall performance. On average, this method put the native structure around 4.5 standard deviations below the population mean, making it easy to see. Some problematic cases were observed, like the sets with the PDB codes 1G1C, 1MLA, and 1R69, in which the native structure was put less than 2 standard deviations below the mean. The PM6 and PM6-DH+ methods were

---

Table 1. Characterization of the decoy sets used in the study.

| Set index | PDB  | Description                              | Type    | Size |
|-----------|------|------------------------------------------|---------|------|
| 1         | 1ABV | F1F0-ATP Synthase delta subunit n-term    | NMR     | 528  |
| 2         | 1AF7 | Chemotaxis receptor CheR                 | X-Ray   | 528  |
| 3         | 1AH9 | IF1 translational initiation factor       | NMR     | 511  |
| 4         | 1AOY | Arginine repressor n-term                | NMR     | 530  |
| 5         | 1B4B | Arginine repressor oligomerization domain| X-Ray   | 461  |
| 6         | 1BM8 | Mbp1 DNA-Binding domain                   | X-Ray   | 330  |
| 7         | 1CEW | Chicken egg white cystatin               | X-Ray   | 453  |
| 8         | 1CQK | Antibody domain CH3                       | X-Ray   | 285  |
| 9         | 1DCJ | E. coli YHHP                              | NMR     | 526  |
| 10        | 1EGX | ENA-VASP Homology 1 domain                | NMR     | 353  |
| 11        | 1FAD | FADD death domain                         | NMR     | 515  |
| 12        | 1FOS | reduced Mj0307                             | NMR     | 342  |
| 13        | 1G1C | II1 Domain from Titin                     | X-Ray   | 306  |
| 14        | 1GIX | Lipoyl Domain from P64K                   | NMR     | 526  |
| 15        | 1GPT | Gamma 1-H Thionin                         | NMR     | 470  |
| 16        | 1ITP | P. ostreatus protease A inhibitor 1        | NMR     | 527  |
| 17        | 1KJS | C5a C-terminus                            | NMR     | 549  |
| 18        | 1KVI | Reduced form of ATP7A                     | NMR     | 551  |
| 19        | 1MLA | Acyl carrier protein transacylase         | X-Ray   | 336  |
| 20        | 1N0U | Yeast elongation factor 2                 | X-Ray   | 302  |
| 21        | 1NE3 | Ribosomal protein 528E                    | NMR     | 566  |
| 22        | 1OF9 | E. histolytica pore forming toxin          | NMR     | 508  |
| 23        | 1R69 | Phage 434 repressor N-term domain         | X-Ray   | 292  |
| 24        | 1SHF | Human Fyn SH3 domain                      | X-Ray   | 537  |
| 25        | 1SRO | S1 RNA binding domain                     | NMR     | 516  |
| 26        | 1TFI | TFII5 nucleic acid-binding domain          | NMR     | 340  |
| 27        | 1TIF | Translational factor IF3 N-term domain    | X-Ray   | 543  |
| 28        | 1TIG | Translational factor IF3 C-term domain    | X-Ray   | 565  |
| 29        | 1VCC | N-termi fragment of vaccinia virus DNA topoisomerase I | X-Ray   | 552  |
| 30        | 2568 | Cytochrome b562                           | X-Ray   | 507  |
| 31        | 2CR7 | PAH domain of the mouse transcriptional repressor SIN3B | NMR     | 541  |
| 32        | 2F3N | Native Shank SAM domai                    | X-Ray   | 486  |
| 33        | 2PCY | Poplar apoplastocyanin                    | X-Ray   | 436  |
overall worse than the PM7 method with instances where the methods put the native inside the population altogether, as exemplified by the sets with PDB codes 1B4B, 1KJS, and 1CEW. In these sets, both of these methods placed the native structure less than one standard deviation below the mean enthalpy. PM6-D3 was not as good as PM7 but better than the other two because it did not place a native structure at less than one standard deviation below the mean enthalpy for any sets, being comparable to PM7 in this aspect.

In Table 2, we can observe the difference between the enthalpies of the native structure and the best ranking decoy. It is easy to distinguish the sets for which the methods failed to correctly rank the native structure, since for such sets the difference is positive. The PM7 method only failed to detect the native structure for the 1KJS set, as did the other methods. However, only a single decoy had its enthalpy lower than the native structure in the ranking of the PM7 method, whereas the PM6 method had 101 decoys with lower enthalpy than the native and the PM6-DH+ had 51. The PM6 method failed to find the native in the sets with PDB codes 1CEW, 1DCJ, 1TFI, 1SRO, and 1KJS. The PM6-DH+ method was unsuccessful only in the sets with PDB codes 1CEW and 1KJS as was PM6-D3, but by a smaller margin.

We compared the methods according to the difference in the heat of formation attributed to the native structure in a set and the line that is three standard deviations below the mean enthalpy of the set, which is the cutoff line for 3SIEVE assessment. These data are shown in Table 3 and summarized in Figure 3.

In Figure 3, the dots that are above the line represent the native structures that were not detected to be native-like, under 3SIEVE. According to 3SIEVE, some methods did not give the native structure a low enough score to set it apart from the rest of the population but these were also native structures that were incorrectly ranked by the scoring function in the first place.

For the sets where the native conformation ranks worse than a decoy, such decoys were not detected as native-like either, even on removal of the native conformation from the set. This means that even though they had a better score, the decoys were not native-like. For the sets where the native structure was properly detected, the PM7 method only detected the 1AOY set to have a decoy as native-like on the removal of the native structure.

This is an indication that for sets where the native structure was not properly discriminated, the energy function collected all the structures around the mean, either because the sampling process was very localized around a known native conformation or the scoring function itself failed to discriminate geometrically dissimilar structures from one another.

The case of the 1KJS set is intriguing. This protein is not large, but it does have three native disulfide bonds as indicated in its PDB record. Some of these bonds are not present in a portion of the decoys. The decoy that had a lower heat of formation than the native conformation according to the PM7 method is part of the fraction that lacks the disulfide bonds.

Initially, it was suspected that the energetics of the disulfide bonds may have been wrongly parameterized. This happened previously with the PM6 method (as mentioned in the PM7 Ref. [41]), and perhaps this problem was not correctly tackled by the new PM7 scheme. However, this hypothesis was quickly refuted because the 1GPT protein has four disulfide bonds and no method failed to detect the native structure for that protein.

As seen in Figure 4, the decoy is a lot more “spread out” than the native, particularly around the amino-terminus (“B” in Fig. 4) and the carboxyl-terminus (“A” in Fig. 4), where one would expect less constrained and faster torsional degrees of freedom to be located. Therefore, we suspected that the structural rigidity provided by disulfide bond formation may interfere with the entropic term of the system, such that the native structure could be enthalpically unfavorable but favorable when considering free energy as a whole. However, that assumption also predicts that such issues would be even more exacerbated in the 1GPT structure, which is not the case. In order further analyze the matter we decided to evaluate the contributions for dispersion interactions (\(E_{\text{disp}}\)), solvation (\(E_{\text{solv}}\)), hydrogen bonding (\(E_{\text{H-bond}}\)), and the total energy (\(E_{\text{tot}}\)), as outputted by MOPAC.
We have computed each of these terms for the native structure and its best decoy, and subtracted the first from the last to generate the variations $\Delta E_{\text{tot}}$, $\Delta E_{\text{disp}}$, $\Delta E_{\text{solv}}$, and $\Delta E_{\text{h-bond}}$. We did that for the 1KJS and additionally to the sets 1BM8, 1AF7, 1CEW, and 1GJX. These results are summarized in Table 4.

Comparisons across the five decoy sets show that for 1KJS the total energy $E_{\text{tot}}$ of the native structure is over 2000 kcal/mol greater than that of its best decoy. Although the difference in total energy is a significant term, it is noticeable from the table that it is not necessarily very large when compared to the other terms. The 1KJS is clearly a specific case where that happens.

Furthermore, although solvation energy of the best decoy for the 1KJS set is nearly 100 kcal/mol lower than that of the native structure, we also notice that this alone is not enough to make a decoy overall better than the native. We see that the best decoy for the 1AF7 set has solvation energy lower than its native's by over 240 kcal/mol, even so, the native has an enthalpy of formation more than 800 kcal/mol lower.
Table 2. The differences for each method and decoy set between the native structure and the best scoring decoy expressed in kcal/mol, and the Z-score of the native function in the set.

| PDB    | PM7  | PM6  | PM6-DH+ | PM6-D3 | PM7  | PM6  | PM6-DH+ | PM6-D3 |
|--------|------|------|---------|--------|------|------|---------|--------|
| 1ABV   | -3.46| -2.60| -2.49   | -2.42  | -870.48| -536.73| -698.24 | -1284.52|
| 1AF7   | -2.89| -3.52| -2.41   | -2.80  | -861.99| -602.77| -692.42 | -1145.81|
| 1AH9   | -3.75| -3.90| -3.31   | -3.34  | -468.40| -279.93| -415.37 | -601.48 |
| 1AOY   | -7.12| -5.53| -4.91   | -4.77  | -743.22| -576.77| -647.31 | -1127.31|
| 1BBR   | -3.30| -0.06| -0.06   | -5.32  | -936.17| -689.34| -805.49 | -1850.38 |
| 1NEB   | -3.61| -6.60| -6.06   | -5.38  | -1390.26| -958.35| -1322.49| -1816.75|
| 1CEW   | -3.35| -0.11| -0.11   | -1.84  | -250.27| 10.42 | 5.57    | -756.23 |
| 1CQK   | -2.59| -3.62| -3.09   | -2.98  | -1415.23| -1188.58| -1286.27| -1617.10|
| 1DCJ   | -7.82| -3.37| -1.87   | -2.70  | -388.45| 31.73 | -157.13 | -1147.89|
| 1EGX   | -3.98| -5.07| -4.09   | -4.66  | -1131.18| -751.22| -928.31 | -1864.95|
| 1FAD   | -13.36| -2.78| -2.15   | -2.14  | -503.56| -236.91| -291.80 | -751.08 |
| 1FOS   | -4.28| -4.67| -4.38   | -4.71  | -804.76| -711.80| -769.46 | -1360.42|
| 1GIC   | -1.90| -5.38| -3.80   | -4.47  | -1316.76| -1047.60| -1182.52| -1979.27|
| 1GJX   | -4.80| -4.50| -3.78   | -4.34  | -608.52| -411.19| -510.89 | -1165.93|
| 1GPT   | -3.00| -3.80| -3.11   | -3.61  | -457.99| -308.44| -404.03| -1016.37|
| 1NTP   | -4.53| -3.88| -3.23   | -3.39  | -575.25| -376.21| -471.57 | -794.16 |
| 1KJS   | -4.75| -1.85| -0.86   | -1.33  | 35.31 | 274.23 | 186.85 | 549.92 |
| 1KVI   | -8.88| -2.90| -2.32   | -2.41  | -316.05| -168.74| -185.43 | -574.59 |
| 1MLA   | -1.63| -4.19| -2.95   | -3.26  | -810.77| -534.78| -700.02 | -1160.92|
| 1N0U   | -3.36| -5.10| -4.54   | -4.59  | -861.19| -704.79| -767.21 | -1336.60|
| 1NQS   | -4.65| -4.45| -3.96   | -4.25  | -608.33| -471.94| -536.50 | -802.74 |
| 1OF9   | -8.47| -2.66| -1.76   | -1.99  | -361.51| -151.56| -211.93 | -478.35 |
| 1R69   | -1.49| -4.75| -3.98   | -4.36  | -664.23| -462.91| -562.06| -1315.62|
| 15HF   | -2.07| -3.47| -2.65   | -2.86  | -606.21| -365.22| -512.21 | -876.87 |
| 1SRO   | -5.57| -3.13| -1.95   | -2.04  | -341.72| 3.92 | -131.90 | -498.17 |
| 1T1F   | -2.76| -2.63| -1.82   | -2.07  | -225.25| 13.08 | -117.50 | -412.00 |
| 1T1F   | -7.96| -4.28| -3.64   | -3.67  | -770.25| -577.23| -672.78 | -1237.33|
| 1TIG   | -3.19| -5.85| -4.79   | -5.46  | -1154.10| -778.83| -961.78 | -1326.70|
| 1VCC   | -5.00| -3.74| -2.71   | -2.94  | -1096.28| -878.49| -988.63 | -1453.25|
| 2568   | -7.11| -4.44| -3.72   | -4.06  | -1028.09| -738.99| -865.25 | -1457.73|
| 2CR7   | -2.49| -3.43| -2.93   | -2.79  | -688.88| -608.87| -611.30| -1073.85|
| 2F3N   | -3.33| -3.43| -2.76   | -3.12  | -671.03| -450.32| -523.52| -921.50 |
| 2PCY   | -4.07| -5.68| -5.04   | -5.64  | -1149.54| -902.25| -1061.76| -1612.31|
| Mean   | -4.57| -3.80| -3.01   | -3.51  | -       | -       | -       | -       |

The calculations were done for the whole set, prior to 3SIEVE application. Z-score ΔΔH.

We also notice from Table 2 that the native of 1AF7 set has an enthalpy of formation over 800 kcal/mol lower than its best decoy’s, whereas for 1CEW that difference is just a bit over 230 kcal/mol. Nevertheless, their solvation, hydrogen bond, and dispersion terms are very alike, which shows the effect of the total energy in the ranking of the decoys.

The total energy is a sum of the electronic energy, which is always negative, and the core–core repulsion energy, which is always positive. In more packed structures, the core–core repulsion energy is expected to be larger. Native structures are more packed than their decoys. In the case of the 1KJS, this can be seen in Figure 4 but that is readily explained because for every decoy set the COSMO cavity for the native structures have a smaller volume than the cavity for each of their respective decoys, also the native structure has a core–core repulsion energy greater than its best decoy.

More densely packed conformations could in principle be stabilized by the other terms such as noncovalent interactions and solvation, which is what usually happens given that every native structure has a larger core–core repulsion energy than its decoys. In the case of the 1KJS however, it is pretty clear that according to PM7 and COSMO, the best decoy is better solvated in water and its total electronic energy is just more favorable than any noncovalent interactions in the native structure. It is also noticeable that decoy solvation alone should not make it automatically better. Either the native has an electronic energy that is too small or a core–core repulsion that is too large.

Decomposing the total energy we verified that the native structure electronic energy is over 460 thousand eV more negative than its best decoy, but its core–core repulsion energy is nearly 460 thousand eV more positive. Since we know that the total energy is not being increased in favor of the best decoy by the electronic energy, which favors the native structure in this case, we have to assume that the main cause for the large total energy in favor of the best decoy comes from the core–core repulsion in the native structure.

The overall success of PM7 in the detection of native structures when compared to the other Hamiltonians with noncovalent corrections is interesting but it might be fruit of some fortuitous error-cancelation when it comes to intramolecular noncovalent interactions, since it has been shown that PM7 is worse than PM6-D3 when compared with experimental free energy variations of interaction between molecules.[149]
The geometric assessment of the decoy quality can be done based on the fraction of the native contacts $Q(X)$, which is reported to be an important geometrical descriptor of protein structure.\cite{50,51} The calculation of this descriptor is done according to eq. (3).

$$Q(X) = \frac{1}{1 + \exp\left[\beta (Q(X) - Q_0^\beta)\right]}$$

In this equation, $X$ is a conformation. $S$ is the set of all heavy atom pairs $(i, j)$ belonging to residues $R_i$ and $R_j$ such that these are more than three residues from each other in the chain and capacity to discriminate geometrically good decoys from geometrically bad ones.

The geometric considerations of the native structure is an interesting feature for a score function to have. However, during a sampling procedure very rarely (if ever) will the sampling algorithm hit the native structure square on. In a realistic setting, an additional procedure very rarely (if ever) will the sampling algorithm hit the score function to have. However, during a sampling procedure very rarely (if ever) will the sampling algorithm hit the native structure square on. In a realistic setting, an additional procedure very rarely (if ever) will the sampling algorithm hit the score function to have.

### Table 3. The difference in the enthalpies between the native structure in each set and its corresponding $-3s$ line in kcal/mol.

| Set index | PD8 | PM7 | PM6 | PM6-DH | PM6-D3 |
|-----------|-----|-----|-----|--------|--------|
| 1         | 1ABV| -640.22 | -367.10 | -485.42 | -549.40 |
| 2         | 1AF7| -696.29 | -452.45 | -542.02 | -558.10 |
| 3         | 1AH9| -347.40 | -221.64 | -324.56 | -345.20 |
| 4         | 1AOY| -682.83 | -533.54 | -602.15 | -628.18 |
| 5         | 1B4B| -805.92 | -574.63 | -685.57 | -729.69 |
| 6         | 1BM8| -1232.85 | -740.33 | -1002.91 | -1063.74 |
| 7         | 1CEW| -136.92 | 132.85 | 41.95 | 16.00 |
| 8         | 1CQK| -1126.25 | -928.06 | -1017.75 | -1036.97 |
| 9         | 1DCJ| -330.52 | 63.86 | -107.39 | -185.92 |
| 10        | 1EEX| -888.23 | -561.15 | -707.42 | -781.92 |
| 11        | 1FAD| -349.34 | -107.13 | -165.85 | -216.85 |
| 12        | 1FOS| -638.82 | -548.11 | -613.38 | -654.20 |
| 13        | 1G1C| -1093.43 | -828.36 | -971.83 | -1013.32 |
| 14        | 1GJX| -521.26 | -330.04 | -429.44 | -507.72 |
| 15        | 1GPT| -398.76 | -254.49 | -348.38 | -358.55 |
| 16        | 1ITP| -466.09 | -275.57 | -364.59 | -399.67 |
| 17        | 1KJS| 115.74 | 386.01 | 288.40 | 251.22 |
| 18        | 1KVI| -233.03 | -72.77 | -103.04 | -149.53 |
| 19        | 1MLA| -640.99 | -418.98 | -529.86 | -549.64 |
| 20        | 1N0U| -740.10 | -586.03 | -657.36 | -679.04 |
| 21        | 1NEI| -483.07 | -351.66 | -412.52 | -444.47 |
| 22        | 1OF9| -194.97 | -12.46 | -72.92 | -96.18 |
| 23        | 1R69| -503.67 | -335.35 | -425.57 | -462.41 |
| 24        | 1SHF| -498.12 | -317.84 | -405.69 | -423.74 |
| 25        | 1SRO| -240.00 | 46.28 | -56.07 | -98.19 |
| 26        | 1TFI| -142.56 | 39.50 | -57.55 | -100.90 |
| 27        | 1TIF| -588.78 | -416.67 | -489.11 | -525.62 |
| 28        | 1TIG| -1033.25 | -690.70 | -854.00 | -901.25 |
| 29        | 1VCC| -916.64 | -720.58 | -814.29 | -850.23 |
| 30        | 256B| -824.14 | -565.88 | -692.05 | -722.60 |
| 31        | 2CR7| -523.17 | -435.28 | -455.54 | -470.61 |
| 32        | 2F3N| -514.76 | -309.58 | -394.87 | -431.01 |
| 33        | 2PCY| -936.04 | -724.18 | -848.78 | -888.52 |

These values were calculated for the set after 3SIEVE removed the upper outliers.

![Figure 3](image_url)  
**Figure 3.** The distance between the native structure enthalpy and the $-3s$ line in each decoy set, for each method.

### Table 4. The decomposition of the terms for the difference between the enthalpies of formation of the native structure and the best decoy for the 1KJS, 1AF7, 1CEW, 1GJX, and 1BM8 sets, according to PM7.

| Energy terms (kcal mol$^{-1}$) | 1KJS | 1AF7 | 1CEW | 1GJX | 1BM8 |
|-------------------------------|------|------|------|------|------|
| $\Delta H_{tot}$              | 2050.71 | -670.87 | 1155.89 | -428.053 | -52.11 |
| $\Delta H_{disp}$             | -109.29 | -121.57 | -94.04 | -139.91 | -205.22 |
| $\Delta H_{bond}$             | -6.95 | -40.85 | -45.11 | -39.93 | -143.74 |
| $\Delta H_{solv}$             | 94.20 | 242.91 | -298.13 | 430.16 | -307.66 |
the distance between atoms \(i\) and \(j\) in the native conformation is less than \(r^*_{ij} = 4.5\ \text{Å}.

The distance between atom \(i\) and \(j\) in conformation \(X\) is denoted \(r_{ij}(X)\), \(|S|\) denotes the number of elements in set \(S\) and \(\beta\) and \(\lambda\) are numerical parameters with the values 5 and 1.8, respectively.\(^{[50]}\)

We used the shorthand notation \(Q(\text{method})\) to denote the fraction of native contacts for the best ranking decoy conformation according to a method. Also in each set there must be a decoy with the greatest fraction of native contacts, the value of that fraction is the \(Q_0\) of the set. The calculation of the fraction of native contacts was made using the suite of scripts pdbTools.\(^{[52]}\)

The graphs in Figure 5 shows the correlations between the semiempirical methods \(Q(\text{PM7}), Q(\text{PM6}), Q(\text{PM6-}
\text{DH}+), Q(\text{PM6-D3})\) and \(Q_0\) for all sets.

Correlation coefficients for all three methods are very similar and in all of them they were slightly larger than 0.8. This is indicative that the best decoys in each set have approximately the same fraction of native contacts as the ones closest to the native structure, which suggests that all three methods manage to find comparably good conformations.

In the absence of the native structure, 3SIEVE did not detect any decoy which it considered native-like, with the exception of a single decoy in the 1AOY set using the PM7 method. An interesting feature of that particular decoy is that it was 3SIEVE-detected only by the PM7 method but it is also considered the best decoy by the other methods.

Further analysis of the decoy shows that its fraction of native contacts is equal to 0.49, which is unusually high for the decoys tested. Furthermore the decoy is also geometrically close to its native structure, as shown in Figure 6, which is also not necessarily the case for the rest of the decoys studied.

One other such decoy with a high fraction of native contacts is the best 1KVI decoy for the PM7, where the same fraction of native contacts was found, but since the population of decoys for 1KVI has less variation, none of them managed to stray too far from the population, triggering the detection by 3SIEVE.

In addition, we found difficulties that the sampling method\(^{[6-8]}\) might have had when it sampled the conformational space, because in most cases we have that \(Q_0 < 0.5\),

---

**Figure 5.** The correlation between the fraction of native contacts of the decoy with the best fraction of native contacts in each set (\(Q_0\)) and the fraction of the native contacts of the decoy with the lowest energy according to each method. Adjusted \(R^2\) correlation coefficients can be seen for the PM6, PM6-DH+, PM6-D3, and PM7 methods in the upper left corner of subgraphs a, b, and c, respectively. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
meaning that the best decoy in each set ordinarily does not satisfy even half of the native contacts.

Conclusions

In this work, we showed that semiempirical Hamiltonians can be good scoring functions for protein structures. We have compared four semiempirical methods within the same family of Hamiltonians from the PM6/PM7 family of Hamiltonians to assess the possibility of using their enthalpies of formation as scoring functions for proteins.

Enthalpically, the PM7 method was found to be superior to the other three methods, not only in identifying the native structure among the decoys in all but one set, but also in statistically distinguishing the native conformation from the rest of the decoys, simulating the deep well in which the thermodynamic hypothesis predicts the native state to be on the free energy surface. Which is a characteristic of a good score function.

When considering only the decoys, however, all the methods manage to find geometrically sound decoys out of the set of conformations, with the PM6-DH+ method faring marginally better than the others. This is shown by the high correlation between the fraction of native contacts of the best decoys according to the methods and the decoys closest to the native. We should also remark that, as a general rule, no decoy in any set had more than half of its native contacts satisfied and the majority had even less. The sole exception was the 1EGX set, for which Qn = 0.57.

Given that the PM7 method had scoring success compared to the PM6 and PM6-DH+ methods but apparently has a better description of the features one would expect in a good scoring landscape, which enabled it to distinguish native structures from the rest, we concluded that the PM7 method is overall a better score function than the PM6, PM6-D3, and PM6-DH+ methods.

Given the increase in computational power that has occurred over the years, the number of structures that can be adequately treated by semiempirical scoring functions will become increasingly larger. It is our view that this computational power will enable the use of quantum descriptors to deal with large numbers of decoys.

Acknowledgments

We would like to thank the CNPq, CAPES, INAMI, FINEP and PRONEX-FACEPE for funding the research as well as CENAPAD/SP for supplying the computational infrastructure and support without which this work would not be possible.

Keywords: semiempirical methods · scoring function · protein decoy set · tertiary structure · enthalpy of formation · quantum chemistry

How to cite this article: G. A. Urquiza-Carvalho, W. D. Fragoso, G. B. Rocha. J. Comput. Chem. 2016, 37, 1962–1972. DOI: 10.1002/jcc.24415
[27] P. Mikulsik, S. Genheden, K. Wichmann, U. Ryde, J. Comput. Chem. 2012, 33, 1179.
[28] N. D. Yilmazer, M. Korth, J. Phys. Chem. B 2013, 117, 8075.
[29] N. D. Yilmazer, P. Heitel, T. Schwabe, M. Korth, J. Theor. Comput. Chem. 2015, 14, 169.
[30] J. Řežáč, J. Fanfrlík, D. Salahub, P. Hobza, J. Chem. Theory Comput. 2009, 5, 1749.
[31] Z. Bikadi, E. Hazai, J. Cheminf. 2009, 1, 1–16.
[32] P. Fanfrlík, A. K. Bronowska, J. Řežáč, O. Pienosił, J. Konvalinka, P. Hobza, J. Phys. Chem. B 2010, 104, 12666.
[33] B. Temelso, K. A. Alser, A. Gauthier, A. K. Palmer, G. C. Shields, J. Phys. Chem. B 2014, 118, 4514.
[34] J. D. L. Dutra, M. A. M. Filho, G. B. Rocha, R. O. Freire, A. M. Simas, J. P. P. Stewart, J. Chem. Theory Comput. 2013, 9, 3333.
[35] J. P. P. Stewart, J. Mol. Model. 2009, 15, 765.
[36] N. D. Yilmazer, P. Heitel, T. Schwabe, M. Korth, J. Theor. Comput. Chem. 2015, 14, 154001.
[37] J. D. C. Maia, G. A. U. Carvalho, C. P. Mangueira, S. R. Santana, L. A. F. Cabral, G. B. Rocha, J. Chem. Theory Comput. 2012, 8, 3072.
[38] A. M. Wollacott, K. M. Merz, Jr., J. Chem. Theory. Comput. 2007, 3, 1609.
[39] Y. A. Arnautova, Y. N. Vorobjev, J. A. Vila, H. A. Scheraga, Proteins 2009, 77, 38.
[40] T. Lazaridis, M. Karplus, J. Mol. Biol. 1998, 288, 477.
[41] J. P. P. Stewart, J. Mol. Model. 2013, 19, 1.
[42] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, J. Chem. Phys. 2010, 132, 154104.
[43] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, J. Comput. Chem. 2004, 25, 1605.
[44] Scigress Version FJ 2.6, Fujitsu limited, copyright 2008-2016. http://www.fujitsu.com/global/solutions/business-technology/tc/sol/scigress/
[45] A. Klamt, A. Schürmann, J. Chem. Soc. Perkin. Trans. 1993, 2, 799.
[46] J. P. P. Stewart, Int. J. Quantum Chem. 1996, 58, 133.
[47] C. L. Brooks, M. Karplus, B. M. Pettitt, Adv. Chem. Phys. 1988, 71, 1.
[48] M. Karplus, T. Ichiyie, B. M. Pettitt, Biophys. J. 1987, 52, 1083.
[49] R. Sure, S. Grimme, J. Chem. Theory Comput. 2015, 11, 3785.
[50] R. B. Best, G. Hummer, W. A. Eaton, Proc Natl Acad Sci USA 2013, 110, 17874.
[51] E. R. Henry, R. B. Best, W. A. Eaton, Proc Natl Acad Sci USA 2013, 110, 17880.
[52] M. Harms, Pdbtools. Available at: https://github.com/harmslab/pdbtools (accessed on 17 May, 2010).