Journal of Molecular Graphics and Modelling 29 (2010) 240–245

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Theoretical prediction of a peptide binding to major histocompatibility complex II

Sarah Aldulaijan, James A. Platts*

School of Chemistry, Cardiff University, Park Place, Cardiff CF10 3AT, UK

ARTICLE INFO

Article history:
Received 24 March 2010
Received in revised form 27 May 2010
Accepted 28 May 2010
Available online 8 June 2010

Keywords:
Non-covalent interactions
Peptides
Semi-empirical methods
Dispersion

ABSTRACT

Prediction of the binding energy of a peptide implicated in multipole sclerosis to its major histocompatibility complex (MHC) receptor is reported using numerous *ab initio*, density functional (DFT) and semi-empirical theoretical methods. Using the crystalline coordinates taken from the protein databank, two *ab initio* methods are shown to be in good agreement for pairwise interaction of amino acids. These data are then used to benchmark more approximate DFT and semi-empirical approaches, which are shown to have substantial errors. However, in some cases significant improvement is apparent on inclusion of an empirical correction to account for dispersion interactions. Most promising among these cases is RM1, a re-parameterisation of the popular AM1 method for atoms typically found in organic and biological molecules. Together with the dispersion correction, this reproduces *ab initio* data with a mean unsigned error of 1.36 kcal/mol. This approach is used to predict binding for progressively larger model systems, up to binding of the peptide with the entire MHC receptor, and is then applied to multiple snapshots taken from molecular dynamics simulation.

1. Introduction

In order to create and develop new drugs, we have to understand the way that a drug interacts with its receptor in order to affect the biological system in the body's cells. One important concept is understanding the chemical interactions between the drug and the receptor [1–3]. In most cases, the most significant interactions between drugs and their biological receptors are non-covalent interactions [4]. Although, non-covalent interactions are typically weaker than covalent interactions, collectively they exert important influence in many properties of biomacromolecules, for example they are well known to affect the structure of proteins, DNA and RNA [5–8].

Accurate and efficient theoretical description of non-covalent interactions is an intense and ongoing area of research [9–11]. Coupled-cluster methods, especially CCSD(T), is widely considered to be a benchmark for non-covalent interaction energies, however; they are not suitable for large complexes calculation due to unfavourable scaling with molecular size [12–15]. MØller–Plesset second-order perturbation theory (MP2), is applicable to larger systems and successfully describes many non-covalent interactions such as hydrogen bonds. However, it is now recognised as not being appropriate for study of stacking interactions, since it overestimates binding energy in cases where dispersion plays a major role [16–18]. According to Sponer, several of the shortcomings of MP2, including those with stacking interactions, can be successfully overcome by use of relatively small basis sets with expanded, diffuse polarisation functions, such as 6-31G(0.25d) [17].

Density fitted, local MP2 (DF-LMP2) makes use of the local nature of electron correlation to further reduce the computational resources required for MP2 calculations [19]. Importantly for the study of non-covalent interactions, this also effectively eliminates basis set superposition error (BSSE), thereby removing the need for potentially expensive counterpoise corrections [20–22]. Spin-component scaling (SCS) MP2 was introduced by Grimme to systematically improve conventional MP2 by separate scaling of the electron correlation due to singlet and triplet electron pairs [23]. More recently, ourselves and others showed that re-parameterisation of this method specifically for non-covalent interactions drastically reduces errors in binding energies. Spin-component scaled for nucleobases (SCSN) [18] and spin-component scaled for molecular interactions (SCS(MI)) [24] methods gave errors of around 0.3 kcal/mol [18] relative to benchmark CCSD(T) calculations for a set of 22 complexes containing a diverse set of interactions, widely denoted as the "S22 set" [6].

Density functional theory (DFT) is widely used in study of large molecules due to its computational efficiency and general accuracy for a wide range of properties [4,8–10,12,25–27]. For instance, DFT has been shown to be highly suitable for study of hydrogen bonding [15,27]. Unfortunately, this method does not succeed

* Corresponding author. Tel.: +44 2920 874950; fax: +44 2920 874030.
E-mail addresses: platts@cardiff.ac.uk, platts@cf.ac.uk (J.A. Platts).

1093-3263/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2010.05.010

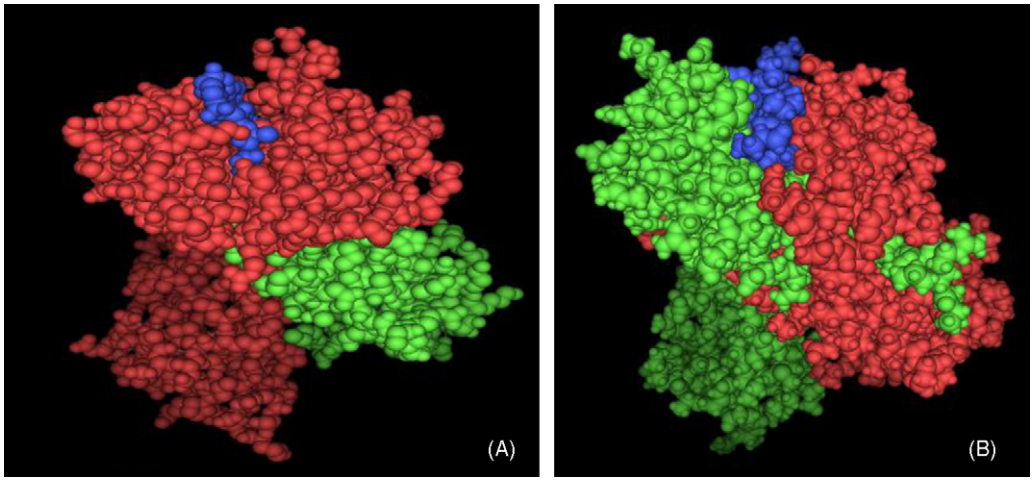

Fig. 1. MHC class I (A) and class II (B). The bound peptide is shown in blue, the receptor chains in red and green.

to describe dispersion interactions, which often play a significant role in non-covalent interactions [28,29]. In order to correct for this fundamental deficiency, DFT methods are often augmented by empirical corrections that mimic the R^{-6} dependence of dispersion [5,7,8,10,14,25,27,30,31]. This method, denoted DFT-D, succeeds to describe the dispersion interaction compared with *ab initio* data, and has been successfully used to study both structures and interaction energies for stacking and hydrogen bonds.

In addition to the DFT-D approach, several DFT methods show promise for description of at least some classes of non-covalent interactions. For example, we recently showed that Becke's Half-and-Half functional (BHandH) performs well for dispersion interactions [16,32,33], providing reliable binding energies and optimised geometries for stacked complexes. Unfortunately, it does not provide good results in studying hydrogen bond interactions, for which the mean unsigned error (MUE) is 5.54 kcal/mol, compared to 0.84 kcal/mol for dispersion-bound complexes. Hybrid meta exchange-correlation functional methods such as M05-2X and M06-2X present accurate calculation for noncovalent interactions especially dispersion interactions [34–36]. The MUE error for dispersion-bound complexes for M06-2X is 0.2 kcal/mol when compared with best estimated values performance for peptides. These methods have been recommended to be used in biochemistry calculations.

Semi-empirical methods are another alternative, their speed and simplicity making them particularly appropriate for study of large systems [8,9,37,38]. Popular semi-empirical methods include AM1 and PM3, but both perform poorly for non-covalent interactions, especially stacking [8]. Analogously to DFT, addition of a dispersion correction term improves performance: AM1-D and PM3-D give errors of 1.1 and 1.2 kcal/mol, respectively, across a wide range of interactions [9]. The OM*x* (orthogonalization model) semi-empirical methods show better results than standard semi-empirical methods such as AM1, especially the OM*x*-D, which include dispersion term. AM1 and OM1-D give errors of 10.67 and 1.08 kcal/mol, respectively, for dispersion interactions [38]. More recent developments in semi-empirical methods include RM1 (Recife Model 1) [37,39,40], which stems from re-optimisation of AM1 parameters for the essential atoms of biological systems C, H, N, O, S, P, F, Cl, Br, and I. MUE values for heat of formation data are 11.15 kcal/mol for AM1, 7.98 kcal/mol for PM3, and 5.77 kcal/mol for RM1. RM1-BH is a further re-parameterisation of RM1 to improve performance specifically for hydrogen bond interactions [41]. Parametric model 6 (PM6) is an extension of and update to PM3 and similar methods, encompassing many more elements within self-consistent set of parameters, which performs well for many classes of compound including hydrogen bonded complexes. PM6-DH is a further development of PM6 to include corrections for the dispersion and H-bond interactions [14,42]. The applicability of semi-empirical methods to large systems is further enhanced by the MOZYME method, using localized molecular orbital instead of the standard SCF procedure, implemented in current versions of MOPAC [43].

Atomistic force fields are widely used in simulation of biological systems by reducing the essentials of systems of interest to simple mathematical forms. Non-covalent interactions are typically treated by a combination of point charges, to account for electrostatics, and Lennard-Jones potentials, for dispersive and repulsive interactions. More than a decade ago, Hobza et al. showed that the force field of Cornell et al. (often referred to as AMBER) best reproduced *ab initio* data for interaction of DNA base pairs [11]. More recently, Paton and Goodman showed that the OPLS-AA force field performs well for binding energy prediction of both hydrogen bonding and dispersion-bound complexes [44].

We have employed many of the methods discussed above to examine in detail the interaction of the immuno-dominant epitope of myelin basic protein (MBP) with Major Histocompatibility Complex (MHC) class II receptor. This epitope, with primary structure Glu-Asn-Pro-Val-Val-His-Phe-Phe-Lys-Asn-Ile-Val-Thr-Pro, is widely used in studies of multiple sclerosis (MS) disease [45]. MHC molecules are an important class of receptor for peptides to study peptide–receptor interactions, and are generally separated into classes I and II. Both have a single peptide binding site, which in class I is made up of a single amino acid chain, whereas in class II the active site is located at the junction between two chains, as shown in Fig. 1 [46,47].

MHC class I includes heavy chain transmembrane glycoproteins α and β2-microglobulin (β2-m), whereas class II consists of two transmembrane glycoprotein chains, α and β [47,48]. The dimensions of the peptide-binding site are around 25 Å long, 10 Å wide and 11 Å deep [48]. Class II can bind with peptides from 13 to 25 residues in length, because the ends of peptide-binding site are open. In contrast, class I binds with peptides from 8 to 11 residues in length [49], since in class I tryptophan-167 and tyrosine-171 of the A pocket, along with tyrosine-84 in the F pocket, act to "close" the binding site: class II does not include such residues [48]. In addition, peptide binding to the MHC, in both classes I and II, occurs because of particular side chain's residues. Many interactions are responsible for MHC-peptide binding, including with peptide side

Table 1
Peptide–protein interactions identified by distance criteria.

Peptide residues	α Residues	β Residues
Glu85	–	–
Asn86	Ser53, Arg50, Phe51	–
Pro87	Ser53, Ala52, Phe51	–
Val88	Ser53	His81
Val89	Ser53, Phe54	Asn82
His90	Phe24	Tyr78, His81, Asn82
Phe91	Gln9, Phe22, Phe54, Gly58, Asn62	Tyr78
Phe92	Gln9, Asn62	Arg13, Phe26, Asp28, Gln70, Ala71, Tyr78
Lys93	Asn62	–
Asn94	Glu11, Asn62, Val65, Asp66	Arg13
Ile95	Val65, Asn69	Trp61, Ile67
Val96	Asn69	Trp61
Thr97	Asn69, Ile72	Asp57, Trp61
Pro98	Arg76	Pro56, Asp57, Tyr60

chain as well as between NH_2 and CHO in the main chains of MHC and peptide [48,49].

2. Methods

The X-ray crystallographic coordinates contained in PDB entry 1YMM, corresponding to the study of Hahn et al. [45] were obtained from the Protein Data Bank [50]. These were loaded into the Molecular Operating Environment (MOE) software, and protonated according to typical protonation states. All hydrogen positions were optimised using the AMBER94 forcefield, with heavy atoms fixed at their X-ray positions. Particular attention was paid to histidine residues, for which all possible protonation states were checked with both OPLSAA and AMBER94. This analysis revealed a clear preference for neutral histidines in all cases, in agreement with Wucherpfenning who stated that LYS93 is the only charged residue on this peptide [51]. Truncation of the PDB coordinates to individual amino acids resulted in neutral species, to avoid charge–charge terms dominating binding energies.

Ab initio calculations were performed using the MOLPRO package of programs [52]. DF-LMP2 calculations and SCSN scaling employed the aug-cc-pVTZ orbital and fitting basis sets [18,53,54]. The Gaussian03 suite of programmes was used to calculate the interaction energies for AM1, PM3, BHandH and MP2/6-31G(0.25d) [55]. The MOPAC programme was used to carry out PM6, RM1 and RM1BH calculations. For larger systems with the RM1 method, we used MOZYME keyword to accelerate the calculations [56]. MOE was used for OPLS-AA and AMBER94 calculations. For PM3-D and AM1-D methods, we used optimised parameters for H, C, N, and O reported by McNamara and Hillier [9]. Dispersion corrections were calculated following the procedure set out by Grimme [57].

3. Results and discussion

MOE was used to obtain each individual pairwise interaction between amino acids in the complex, based on distance criteria between peptide and any atom of receptor residue, as defined in the "ligand interactions" procedure used in MOE [58]. A total of 49 interactions were identified by these criteria, and are listed in Table 1.

A subset of nine interactions, Val88-Ser53α, Phe91-Phe54α, Ile95-Ile67β, Thr97-Asp57β, Pro98-Pro56 β, Pro98-Asp57β, Pro98-Tyr60β, Asn86-Ser53α and Val88-His81β, were selected for study using DF-LMP2 and SCSN methods. Selections were made to cover a range of interaction energies and types. These data were used as a benchmark to test the performance of faster, more approximate methods. As shown in Table 2, MP2/6-31G(0.25d) gives reliable results, with MUE of 0.92 kcal/mol when compared with SCSN. Similar performance (MUE = 0.84 kcal/mol) was found when using SCS(MI) as the benchmark method (see Supporting Information).

Table 2
The MUE (mean unsigned error), MSE (mean signed error), MAX and MIN error for several methods compared with SCSN for seven pairwise amino acid interactions (kcal/mol).

	MUE	MSE	MAX[a]	MIN[a]
PM6-D	7.02	−7.02	−3.22	−13.09
OPLS-AA	6	−4.73	5.32	−17.07
PM3-D	4.22	−3.49	2.13	−8.82
AM1-D	4.17	−1.78	6.64	−6.44
AM1	4.07	1.47	15.96	−4.87
RM1-BH	3.32	−0.12	8.51	−8.1
BHandH	3.19	3.19	8.07	0.41
RM1-D	3.16	−2.48	3.05	−5.59
RM1	2.82	1.38	4.74	−3.36
PM3	2.11	−0.21	2.52	−4.6
PM6	1.81	−1.39	1.83	−6.97
MP2/6-31G(0.25d)	0.92	−0.92	−0.01	−2.11

[a] Errors defined as $E_{MP2} - E_{Method}$, and hence are positive for overbinding relative to MP2, negative for underbinding.

Table 2 also contains data for several more approximate methods. All are considerably worse than MP2/6-31G(0.25d) for these data, although some semi-empirical methods such as PM6, PM3 and RM1 show some promise. Slightly surprisingly, inclusion of dispersion correction (using the default parameters from Ref. [56]) actually makes predictions worse in all cases: this aspect will be discussed in more detail below. For these data at least, OLPS-AA does not appear to be a suitable method to predict interaction energy. On the basis of these results, MP2/6-31G(0.25d) was selected as the most appropriate method to use as a benchmark for all pairwise interactions and for larger systems.

Following this test, all 49 pairwise interactions between amino acids were calculated using several methods: results are summarised in Table 3, using MP2/6-31G(0.25d) as a benchmark, with full details reported in Supporting Information. This data shows that OPLS-AA force field method gave good agreement with MP2 for several interactions, but failed for several others such as PRO98-ASP57, for which the interaction energy is −4.91 kcal/mol with MP2/6-31G(0.25d) and +11.25 kcal/mol according to OPLS-AA. The overall MUE error compared with MP2/6-31G(0.25d) is

Table 3
MUE, MSE, MAX, and MIN relative to MP2/6-31G(0.25d) for all 49 amino acid interactions identified by MOE (kcal/mol).

	MUE	MSE	MAX	MIN
OPLS-AA[a]	2.08	−0.46	5.3	−14.16
PM3	1.97	−1.31	3.21	−8.54
AM1	2.15	−1.74	4.77	−12.96
PM3-D	2.33	1.74	9.51	−4.77
AM1-D	2.33	1.31	7.54	−9.18
RM1BH	2.27	−0.81	14.96	−6.62
PM6	1.68	0.41	15.95	−4.07
RM1	2.17	−1.51	7.29	−6.08
RM1-D	2.13	1.92	8.49	−4.4
RM1-D(0.7)	1.36	−0.14	7.71	−5.41
BHandH	2.39	2.12	15.17	−2.42
PM6-D	3.48	3.46	23.49	−0.36

[a] Cut-off for electrostatic interaction energy of 10 Å employed. Without this cut-off, interaction energies differ by up to 0.17 kcal/mol, MUE identical at the precision shown.

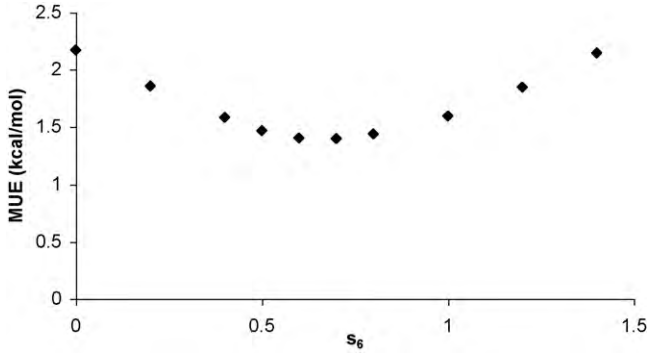

Fig. 2. Response of RM1-D error as a function of global scaling parameter s_6.

2.08 kcal/mol, which although rather smaller than that reported in Table 2 (for fewer interactions) is rather greater than the 1 kcal/mol generally accepted as "chemical accuracy".

BHandH results in overestimated interaction energies, for instance ΔE for PHE92-GLN9 is −12.3 kcal/mol using BHandH and just −5.01 kcal/mol with MP2/6-31G(0.25d). The overall MUE error comparing with MP2/6-31G(0.25) calculations is 2.39 kcal/mol, again rather larger than required for our purposes. Semi-empirical methods PM3 and AM1 give similar overall errors, with MUE of 1.97 kca/mol for PM3 and 2.15 kcal/mol for AM1, and more often than not underestimate interaction energies. Similar performance is found for re-parameterisations RM1 and RM1-BH, whereas PM6 shows a slight improvement over other related methods.

Including a dispersion correction, and modifying parameters according to McNamara and Hillier [9], does not improve performance, with MUE of 2.33 kcal/mol for both AM1-D and PM3-D. To analyse these data in more detail, errors for hydrogen bonded, dispersion bound, and charged interactions were calculated separately, and found to be 1.92, 2.41 and 2.82 kcal/mol for PM3-D, and 2.36, 2.23, and 3.07 kcal/mol for AM1-D, respectively. Notably, Table 3 shows that the mean signed error (MSE) changes sign on addition of dispersion, suggesting that the default dispersion correction overcompensates for the shortcomings of the underlying methods in these cases. The form of the dispersion correction given by Grimme is shown in Eq. (1):

$$E_{disp} = -s_6 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{C_6^{ij}}{R_{ij}^6} f_{dmp}(R_{ij}) \quad (1)$$

where s_6 is a global scaling factor that must be optimised for the method to be corrected. Tables 2 and 3 employed the default scaling factor of 1.4, as in the initial reports of AM1-D and PM3-D method.

Table 4
Stabilisation energies due to each residue in peptide (kcal/mol).

	Sum of pairwise		Direct	Many body
	MP2	RM1-D	RM1-D	
Glu85	–	–	–	–
Asn86	+18.44	+11.51	+14.20	+2.69
Pro87	−3.67	−3.30	−3.21	+0.09
Val88	−4.77	−4.05	−2.09	+1.96
Val89	−4.56	−3.41	−3.52	−0.11
His90	−7.59	−4.77	−3.76	+1.01
Phe91	−6.52	−8.05	−8.36	−0.31
Phe92	−21.93	−13.62	−17.85	−4.23
Lys93	−6.12	−5.86	−5.86	0.00
Asn94	−15.71	−21.08	−16.77	+4.33
Ile95	−1.73	−2.86	−1.60	+1.26
Val96	−6.70	−3.93	−3.84	+0.09
Thr97	+2.22	−2.87	−5.36	−2.49
Pro98	+1.97	+3.28	+0.70	−2.58

Table 5
Interaction energies of larger models (kcal/mol).

Dipeptides	
Glu85-Asn86	10.62
Pro87-Val88	−6.34
Val89-His90	−7.45
Phe91-Phe92	−17.88
Lys93-Asn94	−58.84
Ile95-Val96	−7.63
Thr97-Pro98	1.72
Sum	−85.79
Heptapeptides	
Glu85-Phe91	−20.16
Phe92-Pro98	−81.83
Sum	−101.99

Varying the s_6 parameter, the optimal combination of method and global scaling was located to be RM1 with $s_6 = 0.7$, for which an overall MUE of 1.36 kcal/mol was obtained. The response of this overall error to the value of s_6 is shown in Fig. 2. Repeating this procedure with other semi-empirical methods gave similar curves but slightly higher MUE values, with the exception of PM6 for which the optimal value of s_6 was zero, *i.e.* any attempt to improve on predictions by adding a dispersion term actually gave worse performance.

Previous studies [46,51,59] indicate that the most important residues for interaction of this MBP epitope with MHC-II are the central residues Val88 to Asn94. In general, our calculations are in agreement with these findings, indicating that most stabilisation of the complex stems from these residues' interactions. Table 4 reports the sum of individual pairwise interaction energies from both MP2/6-31G(0.25d) and RM1-D with $s_6 = 0.7$. Both methods show that Phe92 and Asn94 are particularly strongly stabilising, with significant further contributions from all except the terminal residues. Table 4 also reports interaction energies of each peptide residue with *all* receptor residues identified as interacting, calculated with RM1-D. These data show very similar trends to those from pairwise data, with Phe92 and Asn94 considerably more stabilising than other residues, and terminal Asn86 and Pro98 destabilising the complex. The difference between pairwise and direct interaction energies is the many-body term: no clear trend is apparent in this data, but the size of this term in some cases, *e.g.* more than 4 kcal/mol for Phe92, suggests that conclusions from pairwise calculations should be treated with caution. Summing either pairwise or direct interaction energies gives estimates of the overall stabilisation of the complex of −59.0 and −57.3 kcal/mol, respectively.

The speed of semi-empirical methods such as RM1 allows us to examine the interaction energies of larger models of the peptide than single amino acids. The peptide was cut into seven dipeptides, as well as two heptapeptides, and the interaction energy of the peptide with all receptor residues identified in Table 1 calculated, as shown in Table 5. These data further indicate that stabilisation comes mainly from the central residues, with the dipeptide Lys93-Asn94 particularly strongly bound. Table 5 also shows that much greater binding results from the heptapeptide Phe92-Pro98 than from Glu85-Phe91, perhaps unsurprisingly given that this contains both of the most strongly bound single residues.

These data also show that the estimated overall stabilisation of the complex increases as the size of the model peptides increase. Considering only single amino acids, RM1-D estimate is −57.3 kcal/mol, rising to −85.8 kcal/mol from dipeptides and to −102.0 kcal/mol from heptapeptides. Calculating all 14 residues as one molecule with all 29 receptor residues gives an interaction energy of −110.0 kcal/mol, *i.e.* larger still. From these data it seems clear that cutting the peptide into individual amino acids causes significant error in prediction of overall binding energy. The size

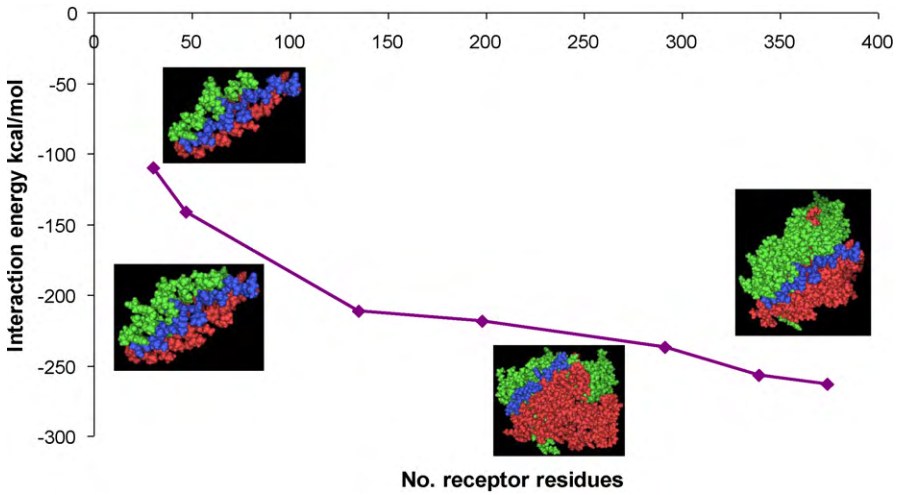

Fig. 3. The interaction energies for large systems (kcal/mol).

Table 6
Interaction energy of peptide with progressively larger models of receptor (kcal/mol).

No. amino acids in receptor model	Total charge (e)	ΔE
29	0	−110.02
47	−1	−141.87
135	−6	−211.13
198	−9	−218.63
291	−9	−236.87
325	−12	−256.49
360 (all)	−13	−263.10

of the receptor was then increased by progressively adding more amino acids, using a 4.5 Å distance cutoff repeatedly until all receptor residues were included. Table 6 shows that there is a large effect of the size of the model on the calculated interaction between peptide and receptor, with the interaction energy of the peptide with the entire receptor calculated at −263.1 kcal/mol. Fig. 3 shows how interaction energy varies with model size, demonstrating that even with almost 300 receptor residues, or 80% of the entire protein, convergence is not reached. It is notable that the total charge on the peptide–protein complex changes from neutrality to −13 as the size of the receptor model increases, suggesting that long-range electrostatic forces may be playing a significant role here.

It is well known that both peptide and receptor are flexible systems, such that calculations on the single static X-ray structure cannot represent the true nature of this interaction. We have therefore calculated interaction energies for ten low-energy snapshots from molecular dynamics (average RMSD Cα = 1.09 Å), taken from Ref. [60], in order to study the effect of the motion in our peptide–receptor interaction energy (Table 7 and Fig. 4). Interaction energies were calculated using OPLS-AA and RM1-D using the entire receptor. These data show that there is a large effect of the motion in our peptide–receptor interaction energy, despite the relatively small RMSD across the snapshots. Detailed analysis of ligand–receptor interactions of each snapshot does not reveal any clear origin for these trends. For instance, snapshot 8 is less strongly bound while snapshot 6 is more strongly bound than average, but snapshot 6 actually has fewer 'native' interactions than snapshot 8 (see Supporting Information). Thus, it seems that the origin of the variation in binding energy is not simple, and cannot be assigned to any single interaction. Slightly surprisingly, in the light of the above results, OPLS-AA shows very similar behaviour to RM1-D results, with a difference in means of 6.9 kcal/mol, and very similar trends in interaction energy across snapshots. Further work is underway

Table 7
Interaction energies for ten molecular dynamics snapshots (kcal/mol).

Snapshot	OPLSAA	RM1-D 40%
1	−175.19	−159.97
2	−192.53	−193.30
3	−177.48	−167.09
4	−157.17	−195.14
5	−193.89	−208.91
6	−194.36	−211.52
7	−178.60	−174.42
8	−121.08	−137.40
9	−175.37	−186.87
10	−150.88	−150.81
Mean	−171.66	−178.54
Standard deviation	22.98	24.78

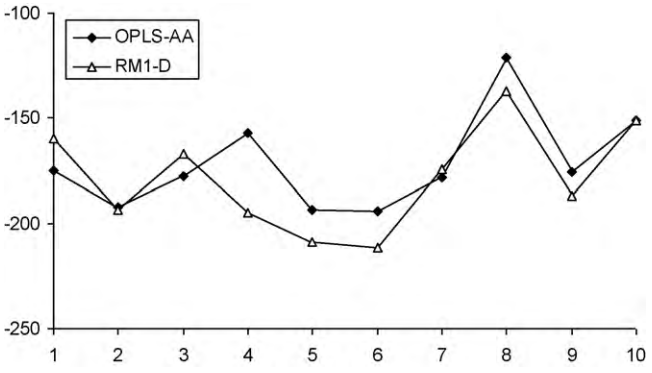

Fig. 4. Interaction energies for ten molecular dynamics snapshots (kcal/mol).

to determine whether longer dynamical simulation and/or more snapshots are required to properly model peptide–protein binding, and results will be reported in a subsequent publication.

4. Conclusions

We have tested several approximate methods against correlated *ab initio* calculations for their ability to predict the energy of interaction between amino acids, focussing on the interaction of a peptide implicated in multiple sclerosis with its biological MHC receptor. We find that the semi-empirical RM1 approach with additional correction for dispersion effects gives the best reproduction of *ab initio* data, with a mean unsigned error of a little more than 1 kcal/mol

over almost 50 interactions after optimisation of the global scaling factor s_6. Performance is similar for several other parameterisations of semi-empirical theory, with RM1 chosen for its slightly better results. The atomistic forcefield OLPS-AA also shows promise, with a mean error of slightly more than 2 kcal/mol.

The computational efficiency of this approach, especially when coupled with the MOZYME method, means that study of larger systems than pairs of amino acids is feasible. Many body effects are significant in some cases, although estimates of complex stabilisation from pairwise interactions and from larger calculations are similar. Increasing the size of model systems used to represent the bound peptide from single amino acids to dipeptides and heptapeptides increases predicting interaction energy, as does expanding the number of amino acids used to model the receptor. We also show that the interaction energy varies significantly over 10 snapshots taken from a previous molecular dynamics study of this complex, confirming that predictions from a single static structure are unlikely to be representative.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2010.05.010.

References

[1] E. Mantzourani, D. Laimou, M.T. Matsoukas, T. Tselios, Anti-Inflamm. Anti-Allergy Agents Med. Chem. 7 (2008) 294–306.
[2] Y. Zhao, D. Truhlar, J. Chem. Theor. Comput. 3 (2007) 289–300.
[3] E.A. Meyer, R.K. Castellano, D. Francois, Angew. Chem. 42 (2003) 1210–1250.
[4] J. Cerny, P. Hobza, Phys. Chem. Chem. Phys. 9 (2007) 5291–5303.
[5] S. Grimme, J. Comput. Chem. 25 (2004) 1463–1473.
[6] P. Jurecka, J. Sponer, J. Cerny, P. Hobza, Phys. Chem. Chem. Phys. 8 (2006) 1985–1993.
[7] P. Jurecka, J. Cerny, P. Hobza, D. Salahub, J. Comput. Chem. 28 (2006) 555–569.
[8] M. Eistner, P. Hobza, T. Frauenheim, S. Suhai, E. Kaxiras, J. Chem. Phys. 114 (2001) 5149–5155.
[9] J.P. McNamara, I.H. Hillier, Phys. Chem. Chem. Phys. 9 (2007) 2362–2370.
[10] R. Sharma, J.P. McNamara, R.K. Raju, M.A. Vincent, I.H. Hillier, C.A. Morgado, Phys. Chem. Chem. Phys. 10 (2008) 2767–2774.
[11] P. Hobza, M. Kabelac, J. Sponer, P. Mejzlik, J.J. Vondrasek, Comput. Chem. 18 (1997) 1136–1150.
[12] K. Berka, R. Laskowski, K.E. Riley, P. Hobza, J. Vondrasek, J. Chem. Theor. Comput. 5 (2009) 982–992.
[13] J.G. Hill, J.A. Platts, H.-J. Werner, Phys. Chem. Chem. Phys. 8 (2006) 4072–4078.
[14] J. Rezac, J. Fanfrlik, D. Salahub, P.J. Hobza, Chem. Theor. Comput. 5 (2009) 1749–1760.
[15] J. Antony, S. Grimme, Phys. Chem. Chem. Phys 8 (2006) 5287–5293.
[16] K. Gkionis, J.G. Hill, S.P. Oldfield, J.A. Platts, J. Mol. Model. 15 (2009) 1051–1060.
[17] J. Sponer, K.E. Riley, P. Hobza, Phys. Chem. Chem. Phys. 10 (2008) 2595–2610.
[18] J.G. Hill, J.A. Platts, J. Chem. Theor. Comput. 3 (2007) 80–85.
[19] H.-J. Werner, P.J. Knowles, F.R. Manby, J. Chem. Phys. 118 (2003) 8149–8160.
[20] P. Pulay, Chem. Phys. Lett. 100 (1983) 151–154.
[21] S. Saebo, P. Pulay, Ann. Rev. Phys. Chem. 44 (1993) 213–236.
[22] C. Hampel, H.-J. Werner, J. Chem. Phys. 104 (1996) 6286–6297.
[23] S. Grimme, J. Chem. Phys. 118 (2003) 9095–9102.
[24] R.A. Distasio, M. Head-Gordon, Mol. Phys. 105 (2007) 1073–1083.
[25] Q. Wu, W. Yang, J. Chem. Phys. 116 (2002) 515–524.
[26] X. Wu, M.C. Vargas, S. Nayak, V. Lotrich, G. Scoles, J. Chem. Phys. 115 (2001) 8748–8757.
[27] U. Zimmerli, M. Parrinello, P. Koumoutsakos, J. Chem. Phys. 120 (2004) 2693–2699.
[28] J. Vondrasek, L. Bendova, V. Klusak, P. Hobza, J. Am. Chem. Soc. 127 (2005) 8232–18232.
[29] M. Tateno, Y. Hagiwara, J. Phys.: Condens. Matter 21 (2009) 243105.
[30] R. Ahlrichs, R. Penco, G. Scoles, Chem. Phys. 19 (1977) 119–130.
[31] C.A. Morgado, J.P. McNamara, I.H. Hillier, N.A. Burton, M.A. Vincent, J. Chem. Theor. Comput. 3 (2007) 1656–1664.
[32] A.D. Becke, J. Chem. Phys. 98 (1993) 1372–1377.
[33] A. Robertazzi, J.A. Platts, Chem. Eur. J. 12 (2006) 5747–5756.
[34] Y. Zhao, N.E. Schultz, D.G. Truhlar, J. Chem. Theor. Comput. 2 (2006) 364–382.
[35] Y. Zhao, D.G. Truhlar, J. Chem. Theor. Comput. 4 (2008) 1849–1868.
[36] K.E. Riley, M. Pitonak, J. Cerny, P. Hobza, J. Chem. Theor. Comput. 6 (2010) 66–80.
[37] G.B. Rocha, R.O. Freire, A.M. Simas, J.J.P. Stewart, J. Comput. Chem. 27 (2006) 1101–1111.
[38] T. Tuttle, W. Thiel, Phys. Chem. Chem. Phys. 10 (2008) 2159–2166.
[39] J.J.P. Stewart, J. Mol. Model. 13 (2007) 1173–1213.
[40] T. Puzyn, N. Suzuki, M. Heranczyk, J. Rak, J. Chem. Inf. Model. 48 (2008) 1174–1180.
[41] F. Feng, H. Wang, W.-H. Fang, J.-G. Yu, J. Theor. Comput. Chem. 8 (2009) 691–711.
[42] M. Korth, M. Pitonak, J. Rezac, P. Hobza, J. Chem. Theor. Comput. 6 (2010) 344–352.
[43] J.J.P. Stewart, MOPAC 2009. Computational Chemistry, Colorado Springs, CO 2008, available at: http://OpenMOPAC.net.
[44] R.S. Paton, J.M. Goodman, J. Chem. Inf. Model. 49 (2009) 944–955.
[45] M. Hahn, M.J. Nicholson, J. Pyrdol, K.W. Wucherpfenning, Nat. Immun. 6 (2005) 490–496.
[46] E.D. Mantzourani, T.M. Mavromoustakos, J.A. Platts, J.M. Matsoukas, T.V. Tselios, Curr. Med. Chem. 12 (2005) 1521–1535.
[47] P.A. Wearsch, P. Cresswell, Curr. Opin. Cell Biol. 20 (2008) 624–631.
[48] L.D. Barber, P. Parham, Ann. Rev. Cell Biol. 9 (1993) 163–206.
[49] J. Yague, A. Marina, J. Vazquez, J.A.L. Castro, J. Biol. Chem. 276 (2001) 43699–43707.
[50] http://www.rcsb.org/pdb.
[51] K.W. Wucherpfenning, A. Sette, S. Southwood, C. Oseroff, M. Matsui, J. Strominger, D.A. Hafler, J. Exp. Med. 179 (1994) 279–290.
[52] H.-J.Werner, P. J. Knowles, R. Lindh, F. R. Manby, M. Schutz, P. Celani, T. Korona, A. Mitrushenkov, G. Rauhut, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, G. Hetzer, T. Hrenar, G. Knizia, C. K¨oppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklass, P. Palmieri, K. Pfluger, R. Pitzer, M. Reiher, U. Schumann, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, M.Wang, and A. Wolf, MOLPRO, version 2008.2, a package of ab initio programs, see http://www.molpro.net.
[53] R.A. Kendall, T.H. Dunning, J. Chem. Phys. 96 (1992) 6796–6806.
[54] F. Weigend, A. Kohn, C. Hattig, J. Chem. Phys. 116 (2002) 3175–3183.
[55] Gaussian 03, Revision C.02, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Wallingford CT, 2004.
[56] J.J.P. Stewart, J. Mol. Model. 15 (2009) 765–805.
[57] S. Grimme, J. Comput. Chem. 27 (2006) 1787–1799.
[58] http://www.chemcomp.com/journal/cstat.htm.
[59] S. Hausmann, M. Martin, L. Gauthier, K. Wucherpfenning, J. Immunol. (1999) 338–344.
[60] E.D. Mantzourani, T.M. Mavromoustakos, J.A. Platts, A. Brancale, T.V. Tselios, J. Mol. Graph. Model. 26 (2007) 471–481.