Supplementary Information for "A generalized many-body expansion and a unified view of fragmentation methods in electronic structure theory"

Ryan M. Richard and John M. Herbert*

Contents

*herbert@chemistry.ohio-state.edu

Here, we wish to expand upon how various existing "one-step" approaches may be decomposed into their various elements. A great bit of detail is left out in the interest of brevity and the interested reader should consult the respective cited works for full and exact details; all nomenclature follows that established in the text. This section is by no means a comprehensive survey of existing fragment approaches, but hopefully illustrates a general classification system. Keep in mind for the coming discussion that the many-body truncation order is part of the fragmentation method and would be represented by a value in parenthesis after the fragmentation method's name, as in XYZ(n). Because extending a method from order two to three (or any such example) presents no real conceptual difference beyond what was explained in the paper, we chose to ignore developments that extend previous approaches to higher orders in the many-body expansion (MBE).

1 ONIOM-like schemes

First, we elaborate upon the fourth element of a fragment approach—the number of layers—since we did not do so in the text. The number of layers is analogous to the number of layers in an ONIOM calculation.[1] In the context of fragment-based approaches, the idea of multilayered calculations has been discussed by Mayhall and Raghavachari[2] and also by Tschumper.[3] One can imagine performing some fragment-based calculation on the entire system and thus obtaining some energy, E_{low}, that approximates the energy of the system. Then one can consider running a "higher level" fragment approach on just part of the overall system (where it is desired to obtain a more accurate description of the electronic structure), and thereby obtaining another energy, E_{high}. We now wish to use the information in both energies, so that all parts of the system not modeled by the higher-level approach are accounted for by E_{low} and all parts included in the higher-level method are not double counted. Clearly,

$$E \approx E_{low} - E^*_{low} + E_{high} \tag{S1}$$

where E^*_{low} represents the energy of the subsystem that was considered using the higher-level fragmentation approach, but evaluated using the lower-level approach. Equation (S1) is simply the energy expression for a two-layer ONIOM calculation, and the generalization to an arbitrary number of layers should be obvious.

Note also that there exists a trivial way of fragmenting a system such that there is one and only one fragment, namely, the entire system, and then the energy corresponding to this fragment is the energy of the whole system at some level of theory. Thus the "number of layers" element also comes into play when one wishes to improve the results of a low-level method, such as Hartree-Fock theory, by using a fragmentation approach at an otherwise inaccessibly high level of theory, say CCSD(T). In this way, many-body induction effects can be handled, say, by a supersystem Hartree-Fock calculation and combined with a low-order MBE of the correlation energy.[4] Various approaches along these lines have been considered recently by Tschumper and co-workers,[3,5–7] and by Dahlke and Truhlar.[8,9] Beran and co-workers[10,11] do this in an even more affordable way by employing a polarizable force field for the low level of theory, combined with two-body MP2 theory for the high level of theory.

From another point of view, ONIOM-like approaches in which one "layer" is the entire supersystem represent a way to *avoid* truncation of the MBE, by summing the MBE to all orders at the "low" level of theory.[10] However, that is not the entire story, as the actual molecules-in-molecules (MIM) equations[2] are more general than that. The low-level system need not be the entire system, but merely an approximation to it, *e.g.*, a truncated MBE. When a truncated MBE is used for the low level and some subsystem is used for the high-level part, the resulting equations will still

be ONIOM-like but are not a way to incorporate higher MBE terms at a lower level of theory. Instead, in the spirit of ONIOM, they will be modeling a particular part of the system at a higher level of theory. Thus, the generalized MBE (GMBE) derived in the paper complements the MIM hierarchy, insofar as one may utilize a truncated GMBE for each energy that appears in the MIM equations.

2 Elemental classification of fragment-based methods

Perhaps the most well-established fragment-based quantum chemistry method is the *fragment molecular orbital* (FMO) approach.[12–14] It relies on user-defined (UD) fragments along with a capping method[12] that we will call "FMO99 capping". In the FMO99 capping approach, one computes localized molecular orbitals (MOs) on related molecules and uses these MOs in the fragment calculations. Embedding is done self-consistently and uses the actual fragment electron densities to compute inter-fragment Coulomb interactions. We will call this "FMO99 embedding method", although it is really just density embedding. Many modifications to the FMO method over the years are simply new embedding methods. For example, the 2002 incarnation[15] introduces two approximations into the FMO99 embedding method, with cutoffs that are managed by a pair of distance parameters. The actual approximations involve separating the inter-fragment interactions into three regions, the first of which (at short range) is treated with density embedding, the second (at intermediate distance) uses density embedding also but exploits a Mulliken approximation for the two-electron integrals, and a third region (for distant fragments) uses self-consistent Mulliken embedding charges. The will refer to this three-region scheme as the "FMO02 embedding method". In 2005, FMO was extended to an arbitrary number of layers.[16]

In 2003, Zhang and Zhang[17] introduced the *molecular fractionation with conjugate caps* (MFCC) approach. The original formulation was applicable only applied to proteins, since the MFCC fragmentation approach defines each fragment as one amino acid. A new capping method was defined in Ref. 17, in which each fragment is capped using the amino acids that appear on either side in the protein sequence. We refer to this as the "MFCC03 capping method".

In 2009, Suárez *et al.*[18] showed that the so-called *kernel energy method* (KEM), introduced previously by Huang *et al.*,[19–22] was simply a normal many-body expansion in which the fragments are rebranded as "kernels". The KEM uses UD fragmentation and a capping method ("KEM capping") that places hydrogen atoms at the positions of the atoms they replace. This method consists of only a single layer without embedding.

Gadre *et al.*[23] introduced the *cardinality-guided molecular tailoring approach* (CG-MTA) in 2006. This method uses a distance-based fragmentation method that is similar to the *generalized energy-based fragmentation* (GEBF) method that is discussed in the paper, except that CG-MTA utilizes two parameters to create the "best" fragments. In other respects, the CG-MTA approach is the same as the KEM.

Another approach is the *electrostatically-embedded many-body* (EE-MB) method of Dahlke and Truhlar.[24] This method uses UD fragmentation without capping, since the method has only been applied to clusters of small molecules, for which the fragments are the molecules themselves. Only one layer of theory is employed.Two embedding methods, called EE-MB-a and EE-MB-b, were described. EE-MB-b embedding employs gas-phase (monomer) Mulliken charges, and is used on our paper also. EE-MB-a embedding charges are computed using the entire cluster instead of a single monomer.

In 2009, Beran[10] introduced the *hybrid many-body interaction approach* (HMBI). This approach uses UD fragmentation, no capping method—as in EE-MB, all systems considered to date consist

of cluster of small molecules—the EE-MB-b embedding method, and two layers of theory. The first layer is a molecular mechanics calculation using a polarizable force field to compute terms through fourth-order in the many-body expansion, whereas the second layer is a two-body electronic structure approach.

Table S1 summarizes the classification of the fragment-based methods that we have just described, as well as those that are discussed in the paper. Perhaps the most appealing feature of our elemental classification is that it is now a straightforward matter to create new approaches—just choose a method from each column of Table S1.

Method	Literature Reference	Fragment Method	Capping Method	Embedding Method	No. of Layers
FMO	12	UD	FMO99	FMO99	1
FMO	15	UD	FMO99	FMO02	1
FMO	16	UD	FMO99	FMO02	n
MFCC	17	MFCC	MFCC03	N/A	1
KEM	19	UD	KEM	N/A	1
CG-MTA	23	CG-MTA	KEM	N/A	1
EE-MB	24	UD	N/A	EE-MB-(a or b)	1
EE-MB-CE	8	UD	N/A	EE-MB-b	2
HMBI	10	UD	N/A	EE-MB-b	2
SFM1	25	SMF1	SMF	N/A	1
SFM2	25	SMF2	SMF	N/A	1
SFM3	25	SMF3	SMF	N/A	1
GEBF	26	GEBF	SMF	XPOL	1

Table S1: Classification of existing fragment-based methods.

References

[1] M. Svensson et al., J. Phys. Chem. **100**, 19357 (1996).

[2] N. J. Mayhall and K. Raghavachari, J. Chem. Theory Comput. **7**, 1336 (2011).

[3] G. S. Tschumper, Chem. Phys. Lett. **427**, 185 (2006).

[4] H. Stoll, Phys. Rev. B **46**, 6700 (1992).

[5] A. M. Elsohly, C. L. Shaw, M. E. Guice, B. D. Smith, and G. S. Tschumper, Mol. Phys. **105**, 2777 (2007).

[6] D. M. Bates, J. R. Smith, T. Janowski, and G. S. Tschumper, J. Chem. Phys. **135**, 044123:1 (2011).

[7] D. M. Bates, J. R. Smith, and G. S. Tschumper, J. Chem. Theory Comput. **7**, 2753 (2011).

[8] E. E. Dahlke and D. G. Truhlar, J. Chem. Theory Comput. **3**, 1342 (2007).

[9] E. E. Dahlke, H. R. Leverentz, and D. G. Truhlar, J. Chem. Theory Comput. **4**, 33 (2008).

[10] G. J. O. Beran, J. Chem. Phys. **130**, 164115:1 (2009).

[11] G. J. O. Beran and K. Nanda, J. Phys. Chem. Lett. **1**, 3480 (2010).

[12] K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, Chem. Phys. Lett. **313**, 701 (1999).

[13] D. G. Fedorov and K. Kitaura, J. Phys. Chem. A **111**, 6904 (2007).

[14] D. G. Fedorov and K. Kitaura, Theoretical background of the fragment molecular orbital (fmo) method and its implementation in gamess, in *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*, edited by D. G. Fedorov and K. Kitaura, chapter 2, pages 5–36, CRC Press, Boca Rotan, FL, 2009.

[15] T. Nakano et al., Chem. Phys. Lett. **351**, 475 (2002).

[16] D. G. Fedorov, T. Ishida, and K. Kitaura, J. Phys. Chem. A **109**, 2638 (2005).

[17] D. W. Zhang and J. Z. H. Zhang, J. Chem. Phys. **119**, 3599 (2003).

[18] E. Suárez, N. Díaz, and D. Suárez, J. Chem. Theory Comput. **5**, 1667 (2009).

[19] L. Huang, L. Massa, and J. Karle, Int. J. Quantum Chem. **103**, 808 (2005).

[20] L. Huang, L. Massa, and J. Karle, Proc. Natl. Acad. Sci. USA **105**, 1849 (2008).

[21] S. N. Weiss, L. Huang, and L. Massa, J. Comput. Chem. **31**, 2889 (2010).

[22] L. Huang, H. J. Bohorquez, C. F. Matta, and L. Massa, Int. J. Quantum Chem. **111**, 4150 (2011).

[23] V. Ganesh, R. K. Dongare, P. Balanarayan, and S. R. Gadre, J. Chem. Phys. **125**, 104109:1 (2006).

[24] E. E. Dahlke and D. G. Truhlar, J. Chem. Theory Comput. **3**, 46 (2007).

[25] V. Deev and M. A. Collins, J. Chem. Phys. **122**, 154102:1 (2005).

[26] W. Li, S. Li, and Y. Jiang, J. Phys. Chem. A **111**, 2193 (2007).

3 Energetic data

		Supersystem	No embedding				Mulliken embedding			
N	Name	Energy	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)
3	C_1	-228.0577651	-228.029432	-228.053579	N/A	N/A	-228.068674	-228.057282	N/A	N/A
	C_3	-228.055649	-228.029583	-228.051717	N/A	N/A	-228.066953	-228.055231	N/A	N/A
	C_{3h}	-228.055004	-228.029987	-228.051638	N/A	N/A	-228.066544	-228.054788	N/A	N/A
4	S_4	-304.086738	-304.037541	-304.074560	-304.085767	N/A	-304.109942	-304.084922	-304.086533	N/A
	C_i	-304.084792	-304.037649	-304.072979	-304.083841	N/A	-304.107367	-304.083046	-304.084598	N/A
	C_4	-304.081904	-304.038027	-304.070583	-304.081023	N/A	-304.104736	-304.080316	-304.081734	N/A
	C_{4h}	-304.081538	-304.038844	-304.071380	-304.080812	N/A	-304.104405	-304.080376	-304.081450	N/A
5	C_1	-380.110278	-380.046422	-380.091217	-380.105234	N/A	-380.143263	-380.107031	-380.109522	N/A
	C_5	-380.105192	-380.047179	-380.087222	-380.100553	N/A	-380.136979	-380.102388	-380.104571	N/A
	C_{5h}	-380.105389	-380.048021	-380.088545	-380.101141	N/A	-380.137286	-380.103118	-380.104947	N/A
6	Prism	-456.139775	-456.055142	-456.124272	-456.1381879	N/A	-456.170940	-456.138584	-456.1393955	N/A
	Cyclic	-456.134384	-456.055672	-456.109417	-456.123840	N/A	-456.176312	-456.130155	-456.133093	N/A
	Cage	-456.137827	-456.055475	-456.122092	-456.136338	N/A	-456.172045	-456.136651	-456.1373918	N/A
	Book1	-456.136202	-456.055448	-456.116022	-456.131244	N/A	-456.174506	-456.133306	-456.135525	N/A
	Book2	-456.135567	-456.055345	-456.116044	-456.129668	N/A	-456.173545	-456.132678	-456.134945	N/A
	Bag	-456.134101	-456.055044	-456.113778	-456.127439	N/A	-456.171873	-456.131212	-456.133136	N/A
	Boat1	-456.132226	-456.055821	-456.108373	-456.122071	N/A	-456.173213	-456.128147	-456.131098	N/A
	Boat2	-456.132612	-456.055873	-456.108757	-456.123901	N/A	-456.173196	-456.128609	-456.1314469	N/A

Table S2: Energies (a.u.) of $(H_2O)_N$ clusters at the HF/6-31G* level of theory. Cluster geometries and nomenclature are taken from Ref. 7.

N	Name	Supersystem Energy	No embedding				Mulliken embedding			
			SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)
7	OptA	-532.166799	-532.063809	-532.145048	-532.161625	N/A	-532.210747	-532.164536	-532.166047	N/A
	OptB	-532.162824	-532.063959	-532.140302	-532.158424	-532.162830	-532.208429	-532.160214	-532.161959	-532.162825
	OptC	-532.159145	-532.064361	-532.132677	-532.149922	-532.159149	-532.207195	-532.154871	-532.157866	-532.159144
	OptD	-532.159608	-532.064558	-532.136154	-532.158208	-532.159612	-532.204964	-532.156425	-532.159206	-532.159614
	PR2	-532.166642	-532.063559	-532.145260	-532.161158	N/A	-532.209673	-532.164641	-532.165862	N/A
	PR3	-532.165788	-532.063925	-532.144332	-532.161497	N/A	-532.209550	-532.163644	-532.165007	N/A
	CA1	-532.162756	-532.064164	-532.139735	-532.156129	N/A	-532.208651	-532.159985	-532.161876	N/A
	CA2	-532.161256	-532.063928	-532.138666	-532.156047	N/A	-532.206450	-532.158600	-532.160496	N/A
	CH1	-532.159690	-532.064320	-532.133224	-532.151212	-532.159693	-532.207455	-532.155402	-532.158407	-532.159691
	CH2	-532.160258	-532.064285	-532.133574	-532.149956	-532.160261	-532.208468	-532.155931	-532.158977	-532.160253
	CH3	-532.159686	-532.064307	-532.133214	-532.151214	-532.159682	-532.207421	-532.155500	-532.158400	-532.159689
8	D_{2d}	-608.201264	-608.071522	-608.174549	-608.209131	N/A	-608.257562	-608.199332	-608.201484	N/A
	S_4	-608.201062	-608.071420	-608.173853	-608.191740	N/A	-608.25795	-608.198774	-608.199757	N/A
	C_i	-608.196388	-608.071479	-608.170392	-608.183607	N/A	-608.250161	-608.193879	-608.194764	N/A
	C_s	-608.194610	-608.070814	-608.168410	-608.182391	N/A	-608.247684	-608.191991	-608.192960	N/A
	C_2	-608.196559	-608.071633	-608.171005	-608.200363	N/A	-608.249902	-608.194100	-608.196520	N/A
	C_1b	-608.194864	-608.071790	-608.1700610	-608.181306	N/A	-608.247343	-608.192558	-608.195644	N/A
	C_1a	-608.194949	-608.072315	-608.171202	-608.200330	N/A	-608.246756	-608.192806	-608.1950133	N/A
	C_1c	-608.194825	-608.072196	-608.170669	-608.183960	N/A	-608.246978	-608.192736	-608.1933220	N/A
	Noncubic	-608.193035	-608.072753	-608.168049	-608.193749	-608.193027	-608.246507	-608.190804	-608.192903	-608.193026
9	D_{2d}Dh	-684.225206	-684.080315	-684.190976	-684.228687	-684.225190	-684.292844	-684.221424	-684.224748	-684.2251989
	S_4Dah1	-684.224489	-684.079975	-684.190187	-684.209139	-684.224483	-684.291884	-684.220542	-684.222544	-684.224475
	S_4DAh2	-684.224439	-684.079987	-684.190116	-684.209235	-684.224425	-684.291921	-684.220500	-684.222498	-684.224424
	S_4DDh 1	-684.224591	-684.080220	-684.190822	-684.211516	-684.224584	-684.291866	-684.220844	-684.222741	-684.224586
	S_4DDh 2	-684.224630	-684.080260	-684.190868	-684.211609	-684.224617	-684.291924	-684.220899	-684.222765	-684.224618
	D_{2d}Dah	-684.224215	-684.080090	-684.191223	-684.228065	-684.224206	-684.290591	-684.220639	-684.223970	-684.224201
	S_4Danh 1	-684.223930	-684.080006	-684.190710	-684.213918	-684.223915	-684.290785	-684.220181	-684.222621	-684.223927
	S_4Danh 2	-684.224047	-684.080043	-684.1908624	-684.215769	-684.224035	-684.290555	-684.220493	-684.222653	-684.224041

Table S3: Continuation of Table S2.

		Supersystem	No embedding				Mulliken embedding			
N	Name	Energy	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)
10	PP1	-760.253433	-760.088565	-760.213839	-760.258115	-760.253409	-760.330813	-760.249047	-760.252886	-760.253524
	PP2	-760.253253	-760.088402	-760.213267	-760.229200	-760.253213	-760.330513	-760.248747	-760.250088	-760.253296
	PP3	-760.252417	-760.088318	-760.214887	-760.250550	-760.252411	-760.328115	-760.248357	-760.251609	-760.252405
	PP4	-760.252195	-760.088136	-760.214103	-760.233184	-760.252207	-760.328294	-760.247735	-760.249929	-760.252146
	PP5	-760.252425	-760.088466	-760.214570	-760.242477	-760.252380	-760.328184	-760.248126	-760.250826	-760.252488
	OB1	-760.252723	-760.089896	-760.217052	-760.258387	-760.252730	-760.324781	-760.249889	-760.252567	-760.252708
	OB2	-760.252433	-760.089784	-760.216356	-760.239131	-760.252423	-760.324868	-760.249076	-760.250678	-760.252457
	OB3	-760.252487	-760.089798	-760.216397	-760.240210	-760.252474	-760.324890	-760.249857	-760.250658	-760.252449
	DP1	-760.250489	-760.088705	-760.211837	-760.251125	-760.250479	-760.325425	-760.245470	-760.249666	-760.250430
	OB4	-760.251782	-760.089363	-760.218187	-760.261135	-760.251784	-760.322442	-760.249802	-760.252008	-760.251796
	OB5	-760.252726	-760.089898	-760.217056	-760.258389	-760.252723	-760.324761	-760.250353	-760.252606	-760.252703
	DP2	-760.250295	-760.088694	-760.211113	-760.233799	-760.250316	-760.325932	-760.245507	-760.248144	-760.250322
	OB6	-760.251554	-760.089562	-760.217863	-760.260824	-760.251544	-760.322602	-760.249231	-760.251803	-760.251594
	OB7	-760.251743	-760.089697	-760.217429	-760.244808	-760.251736	-760.322723	-760.248558	-760.250794	-760.251709
	OB8	-760.251765	-760.089733	-760.217743	-760.254289	-760.251758	-760.322457	-760.249091	-760.251476	-760.25180
	DP3	-760.248206	-760.088465	-760.206926	-760.226664	-760.248212	-760.325953	-760.243145	-760.245584	-760.248201
	DP4	-760.248348	-760.088896	-760.207995	-760.247604	-760.248311	-760.325425	-760.242834	-760.247278	-760.248409
	DP5	-760.248514	-760.088973	-760.208227	-760.232386	-760.248519	-760.325929	-760.243254	-760.246104	-760.248464
	DP6	-760.247821	-760.088395	-760.207535	-760.231086	-760.247837	-760.325201	-760.242064	-760.245903	-760.247874
	OB9	-760.251015	-760.090100	-760.215347	-760.229618	-760.250998	-760.322512	-760.248170	-760.248559	-760.251016
	DP7	-760.248056	-760.088915	-760.206882	-760.228162	-760.248065	-760.326149	-760.242382	-760.245332	-760.248008
	DP8	-760.248120	-760.089035	-760.207921	-760.231795	-760.248152	-760.325473	-760.242486	-760.245697	-760.248090
	OB10	-760.252622	-760.089757	-760.216907	-760.258310	-760.252611	-760.324989	-760.249970	-760.252544	-760.252619
	OB11	-760.250268	-760.089523	-760.216591	-760.244034	-760.250260	-760.321591	-760.247788	-760.248920	-760.250256
	DP10	-760.248112	-760.088989	-760.207891	-760.231717	-760.248152	-760.325338	-760.242600	-760.245675	-760.248140
	DP11	-760.247086	-760.088509	-760.206819	-760.233363	-760.247038	-760.324065	-760.241372	-760.245070	-760.247071
	C_1	-760.246605	-760.089451	-760.207801	-760.225257	-760.246592	-760.320547	-760.241804	-760.244293	-760.246594
	C_2	-760.246693	-760.090078	-760.208381	-760.235091	-760.246745	-760.320956	-760.242173	-760.245001	-760.246630
	C_3	-760.246215	-760.089885	-760.207791	-760.229875	-760.246193	-760.320707	-760.241363	-760.244168	-760.246186

Table S4: Continuation of Table S2.

		Supersystem	No embedding				Mulliken embedding			
N	Name	Energy	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)
3	C_1	-229.335292	-229.309194	-229.331433	N/A	N/A	-229.345920	-229.334910	N/A	N/A
	C_3	-229.333924	-229.309222	-229.330208	N/A	N/A	-229.345416	-229.333593	N/A	N/A
	C_{3h}	-229.332661	-229.309258	-229.329796	N/A	N/A	-229.343901	-229.332558	N/A	N/A
4	S_4	-305.792762	-305.744879	-305.781876	-305.791862	N/A	-305.811169	-305.791007	-305.792538	N/A
	C_i	-305.791159	-305.744928	-305.780578	-305.789880	N/A	-305.809556	-305.789704	-305.790960	N/A
	C_4	-305.788924	-305.745043	-305.778730	-305.787925	N/A	-305.808263	-305.787457	-305.788708	N/A
	C_{4h}	-305.787552	-305.745236	-305.778637	-305.786415	N/A	-305.806467	-305.786119	-305.787130	N/A
5	C_1	-382.244490	-382.180849	-382.227552	-382.240153	N/A	-382.269457	-382.24188	6-382.2439603	N/A
	C_5	-382.239496	-382.181133	-382.224283	-382.2349630	N/A	-382.265233	-382.237432	-382.239181	N/A
	C_{5h}	-382.238589	-382.181378	-382.223975	-382.235050	N/A	-382.264297	-382.236798	-382.238262	N/A
6	Prism	-458.693061	-458.616655	-458.679856	-458.6908632	N/A	-458.723999	-458.691364	-458.692897	N/A
	Cyclic	-458.695851	-458.616987	-458.673967	-458.686786	N/A	-458.726910	-458.692616	-458.694889	N/A
	Cage	-458.693375	-458.616987	-458.678646	-458.692137	N/A	-458.724979	-458.691518	-458.693214	N/A
	Book1	-458.695268	-458.616970	-458.677406	-458.690630	N/A	-458.726292	-458.692519	-458.694808	N/A
	Book2	-458.694759	-458.616914	-458.677652	-458.688874	N/A	-458.726004	-458.692025	-458.694416	N/A
	Bag	-458.693256	-458.616668	-458.675160	-458.687159	N/A	-458.724479	-458.690409	-458.692545	N/A
	Boat1	-458.693742	-458.617047	-458.673295	-458.684486	N/A	-458.724727	-458.690655	-458.692996	N/A
	Boat2	-458.693666	-458.617075	-458.673414	-458.685487	N/A	-458.724126	-458.690517	-458.692815	N/A

Table S5: Energies (a.u.) of $(H_2O)_N$ clusters at the B3LYP/6-31+G(d,2p) level of theory. Cluster geometries and nomenclature are taken from Ref. 7.

N	Name	Supersystem Energy	No embedding				Mulliken embedding			
			SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)
7	OptA	-535.149179	-535.052705	-535.130421	-535.143014	N/A	-535.188233	-535.146292	-535.148786	N/A
	OptB	-535.147204	-535.052778	-535.127344	-535.143175	-535.147451	-535.186011	-535.144180	-535.146517	-535.147221
	OptC	-535.146652	-535.052939	-535.123741	-535.137757	-535.145062	-535.184394	-535.142936	-535.145696	-535.146629
	OptD	-535.145431	-535.053161	-535.124275	-535.145132	-535.144092	-535.182248	-535.142527	-535.145272	-535.145454
	PR2	-535.148532	-535.052520	-535.129379	-535.143088	N/A	-535.187440	-535.145623	-535.148085	N/A
	PR3	-535.148224	-535.052683	-535.129738	-535.142879	N/A	-535.187053	-535.145382	-535.147550	N/A
	CA1	-535.147429	-535.052874	-535.127519	-535.141367	N/A	-535.185907	-535.144431	-535.146705	N/A
	CA2	-535.146054	-535.052708	-535.125810	-535.140940	N/A	-535.184566	-535.142723	-535.145324	N/A
	CH1	-535.146841	-535.052926	-535.123935	-535.138751	-535.145966	-535.184188	-535.143015	-535.145867	-535.146876
	CH2	-535.147620	-535.052915	-535.124639	-535.137619	-535.146398	-535.185217	-535.143832	-535.146722	-535.147656
	CH3	-535.146840	-535.052921	-535.123928	-535.138759	-535.145950	-535.184184	-535.143124	-535.145867	-535.146852
8	D_{2d}	-611.610140	-611.488363	-611.587221	-611.616522	N/A	-611.659482	-611.606291	-611.610454	N/A
	S_4	-611.610460	-611.488295	-611.586809	-611.599807	N/A	-611.659896	-611.606236	-611.609375	N/A
	C_i	-611.605470	-611.488196	-611.582295	-611.592795	N/A	-611.653518	-611.601467	-611.604356	N/A
	C_s	-611.603756	-611.487643	-611.581398	-611.591612	N/A	-611.651549	-611.600264	-611.603173	N/A
	C_2	-611.605584	-611.488292	-611.582857	-611.607154	N/A	-611.652984	-611.601628	-611.605783	N/A
	C_1b	-611.603903	-611.488309	-611.582005	-611.590744	N/A	-611.650800	-611.600473	-611.605171	N/A
	C_1a	-611.603509	-611.488659	-611.582520	-611.607151	N/A	-611.649967	-611.600043	-611.603834	N/A
	C_1c	-611.603272	-611.488585	-611.582085	-611.592376	N/A	-611.650479	-611.599987	-611.602525	N/A
	Noncubic	-611.602819	-611.488905	-611.580399	-611.602829	-611.602393	-611.649236	-611.599171	-611.602436	-611.602749
9	D_{2d}Dh	-688.063118	-687.924291	-688.033604	-688.065605	-688.063301	-688.119367	-688.058033	-688.062863	-688.063190
	S_4Dah1	-688.062471	-687.924096	-688.032409	-688.047994	-688.062414	-688.119215	-688.057349	-688.061149	-688.062488
	S_4Dah2	-688.062474	-687.924088	-688.032373	-688.048023	-688.062762	-688.119191	-688.057458	-688.061015	-688.062497
	S_4DDh 1	-688.0626202	-687.924240	-688.033217	-688.050365	-688.062427	-688.118698	-688.057386	-688.061225	-688.062623
	S_4DDh 2	-688.062462	-687.924267	-688.033423	-688.050452	-688.063396	-688.118836	-688.057590	-688.061352	-688.062707
	D_{2d}Dah	-688.062004	-687.924167	-688.032738	-688.064976	-688.063962	-688.117776	-688.057075	-688.0621641	-688.062012
	S_4Danh 1	-688.062048	-687.924115	-688.032722	-688.052737	-688.061376	-688.118274	-688.057008	-688.061238	-688.062053
	S_4Danh 2	-688.061739	-687.924133	-688.032415	-688.053703	-688.061375	-688.117708	-688.056853	-688.060708	-688.061699

Table S6: Continuation of Table S5.

		Supersystem	No embedding				Mulliken embedding			
N	Name	Energy	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)
10	PP1	-764.518422	-764.359965	-764.483932	-764.521058	-764.521445	-764.582404	-764.512427	-764.518120	-764.518539
	PP2	-764.518183	-764.359904	-764.484003	-764.495150	-764.519487	-764.582246	-764.512117	-764.515606	-764.518326
	PP3	-764.517043	-764.359823	-764.484415	-764.514936	-764.517050	-764.580531	-764.511183	-764.516726	-764.5172929
	PP4	-764.517141	-764.359713	-764.483584	-764.499196	-764.519189	-764.581018	-764.510798	-764.515219	-764.517129
	PP5	-764.517017	-764.359944	-764.483700	-764.507619	-764.518084	-764.580615	-764.510848	-764.515751	-764.517010
	OB1	-764.5140466	-764.360848	-764.482710	-764.518130	-764.515555	-764.577078	-764.509091	-764.514174	-764.514120
	OB2	-764.513982	-764.360804	-764.482882	-764.501048	-764.516076	-764.577442	-764.508555	-764.512743	-764.514010
	OB3	-764.514063	-764.360813	-764.482852	-764.501479	-764.516186	-764.577420	-764.509305	-764.512725	-764.514019
	DP1	-764.514999	-764.359905	-764.481293	-764.515348	-764.516360	-764.577361	-764.508734	-764.514502	-764.515172
	OB4	-764.513564	-764.360458	-764.483333	-764.521513	-764.511884	-764.575149	-764.509338	-764.513968	-764.513636
	OB5	-764.514048	-764.360849	-764.482716	-764.518128	-764.515557	-764.577064	-764.509596	-764.514251	-764.514143
	DP2	-764.514945	-764.359918	-764.480975	-764.499340	-764.516226	-764.578000	-764.509027	-764.513223	-764.514861
	OB6	-764.513211	-764.360625	-764.483093	-764.521193	-764.512987	-764.575616	-764.508658	-764.513758	-764.513233
	OB7	-764.513480	-764.360698	-764.482483	-764.506065	-764.512842	-764.576004	-764.508235	-764.512979	-764.513460
	OB8	-764.513381	-764.360717	-764.482364	-764.515018	-764.511607	-764.575596	-764.508694	-764.513447	-764.513470
	DP3	-764.514881	-764.359867	-764.479342	-764.495418	-764.512858	-764.578167	-764.509073	-764.512945	-764.514838
	DP4	-764.514538	-764.360128	-764.479513	-764.513612	-764.510315	-764.577265	-764.508230	-764.513871	-764.514523
	DP5	-764.514744	-764.360188	-764.480119	-764.499323	-764.513137	-764.577433	-764.508542	-764.512886	-764.514694
	DP6	-764.514784	-764.359816	-764.479419	-764.499726	-764.510997	-764.577962	-764.508071	-764.513523	-764.514815
	OB9	-764.512711	-764.361089	-764.481491	-764.492177	-764.515225	-764.576021	-764.507889	-764.511053	-764.512652
	DP7	-764.514831	-764.360094	-764.479954	-764.495695	-764.514842	-764.578217	-764.508489	-764.512875	-764.514846
	DP8	-764.514302	-764.360219	-764.480211	-764.499126	-764.512907	-764.577091	-764.507806	-764.512441	-764.514281
	OB10	-764.514064	-764.360762	-764.482589	-764.518314	-764.515182	-764.577308	-764.509264	-764.514217	-764.514122
	OB11	-764.512208	-764.360532	-764.482494	-764.505842	-764.513318	-764.574992	-764.507548	-764.511373	-764.512234
	DP10	-764.514228	-764.360207	-764.480127	-764.499111	-764.512921	-764.577003	-764.507890	-764.512386	-764.514165
	DP11	-764.513746	-764.359837	-764.477865	-764.501098	-764.510186	-764.576465	-764.506997	-764.512270	-764.513716
	C_1	-764.511042	-764.360496	-764.476966	-764.491741	-764.510890	-764.573250	-764.505109	-764.509371	-764.511053
	C_2	-764.510924	-764.360797	-764.477477	-764.500007	-764.508225	-764.573115	-764.505468	-764.509421	-764.510920
	C_3	-764.510867	-764.360642	-764.477718	-764.495939	-764.509636	-764.572892	-764.505241	-764.509264	-764.510858

Table S7: Continuation of Table S5.

		Supersystem	No embedding				Mulliken embedding			
N	Name	Energy	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)
3	C_1	0.00	0.00	0.00	N/A	N/A	0.00	0.00	N/A	N/A
	C_3	1.33	-0.09	1.17	N/A	N/A	1.08	1.29	N/A	N/A
	C_{3h}	1.73	-0.35	1.22	N/A	N/A	1.34	1.57	N/A	N/A
4	S_4	0.00	0.00	0.00	0.00	N/A	0.00	0.00	0.00	N/A
	C_i	1.22	-0.07	0.99	1.21	N/A	1.62	1.18	1.21	N/A
	C_4	3.03	-0.30	2.50	2.98	N/A	3.27	2.89	3.01	N/A
	C_{4h}	3.26	-0.82	2.00	3.11	N/A	3.47	2.85	3.19	N/A
5	C_1	0.00	0.00	0.00	0.00	N/A	0.00	0.00	0.00	N/A
	C_5	3.19	-0.48	2.51	2.94	N/A	3.94	2.91	3.11	N/A
	C_{5h}	3.07	-1.00	1.68	2.57	N/A	3.75	2.46	2.87	N/A
6	Prism	0.00	0.00	0.00	0.00	N/A	0.00	0.00	0.00	N/A
	Cyclic	3.38	-0.33	9.32	9.00	N/A	-3.37	5.29	3.95	N/A
	Cage	1.22	-0.21	1.37	1.16	N/A	-0.69	1.21	1.26	N/A
	Book1	2.24	-0.19	5.18	4.36	N/A	-2.24	3.31	2.43	N/A
	Book2	2.64	-0.13	5.16	5.35	N/A	-1.63	3.71	2.79	N/A
	Bag	3.56	0.06	6.59	6.75	N/A	-0.59	4.63	3.93	N/A
	Boat1	4.74	-0.43	9.98	10.11	N/A	-1.43	6.55	5.21	N/A
	Boat2	4.49	-0.46	9.74	8.97	N/A	-1.42	6.26	4.99	N/A

Table S8: Relative energies (kcal/mol) of $(H_2O)_N$ clusters at the HF/6-31G* level of theory. Cluster geometries and nomenclature are taken from Ref. 7.

N	Name	Supersytem Energy	No embedding				Mulliken embedding			
			SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)
7	OptA	0.00	0.00	0.00	0.00	N/A	0.00	0.00	0.00	N/A
	OptB	2.49	-0.09	2.98	2.01	N/A	1.45	2.71	2.57	N/A
	OptC	4.80	-0.35	7.76	7.34	N/A	2.23	6.06	5.13	N/A
	OptD	4.51	-0.47	5.58	2.14	N/A	3.63	5.09	4.29	N/A
	PR2	0.10	0.16	-0.13	0.29	N/A	0.67	-0.07	0.12	N/A
	PR3	0.63	-0.07	0.45	0.08	N/A	0.75	0.56	0.65	N/A
	CA1	2.54	-0.22	3.33	3.45	N/A	1.32	2.86	2.62	N/A
	CA2	3.48	-0.07	4.00	3.50	N/A	2.70	3.72	3.48	N/A
	CH1	4.46	-0.32	7.42	6.53	N/A	2.07	5.73	4.79	N/A
	CH2	4.10	-0.30	7.20	7.32	N/A	1.43	5.40	4.44	N/A
	CH3	4.46	-0.31	7.43	6.53	N/A	2.09	5.67	4.80	N/A
8	D_{2d}	0.00	0.00	0.00	0.00	N/A	0.00	0.00	0.00	N/A
	S_4	0.13	0.06	0.44	10.91	N/A	-0.24	0.35	1.08	N/A
	C_i	3.06	0.03	2.61	16.02	N/A	4.64	3.42	4.22	N/A
	C_s	4.18	0.44	3.85	16.78	N/A	6.20	4.61	5.35	N/A
	C_2	2.95	-0.07	2.22	5.50	N/A	4.81	3.28	3.11	N/A
	C_1b	4.02	-0.17	2.82	17.46	N/A	6.41	4.25	3.66	N/A
	C_1a	3.96	-0.50	2.10	5.52	N/A	6.78	4.10	4.06	N/A
	C_1c	4.04	-0.42	2.43	15.8	N/A	6.64	4.14	5.12	N/A
	Noncubic	5.16	-0.77	4.08	9.65	N/A	6.94	5.35	5.38	N/A
9	D_{2d}Dh	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	S_4Dah 1	0.45	0.21	0.50	12.27	0.44	0.60	0.55	1.38	0.45
	S_4Dah 2	0.48	0.21	0.54	12.21	0.48	0.58	0.58	1.41	0.49
	S_4DDh 1	0.39	0.06	0.10	10.77	0.38	0.61	0.36	1.26	0.38
	S_4DDh 2	0.36	0.03	0.07	10.72	0.36	0.58	0.33	1.24	0.36
	D_{2d}Dah	0.62	0.14	-0.15	0.39	0.62	1.41	0.49	0.49	0.63
	S_4Danh 1	0.80	0.19	0.17	9.27	0.80	1.29	0.78	1.33	0.80
	S_4Danh 2	0.73	0.17	0.07	8.11	0.72	1.44	0.58	1.31	0.73

Table S9: Continuation of Table S8.

		Supersystem	No embedding				Mulliken embedding			
N	Name	Energy	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)
10	PP1	0	0	0	0	0	0	0	0	0
	PP2	0.11	0.1	0.36	18.14	0.12	0.19	0.19	1.76	0.14
	PP3	0.64	0.15	-0.66	4.75	0.63	1.69	0.43	0.8	0.7
	PP4	0.78	0.27	-0.17	15.64	0.75	1.58	0.82	1.86	0.86
	PP5	0.63	0.06	-0.46	9.81	0.65	1.65	0.58	1.29	0.65
	OB1	0.45	-0.84	-2.02	-0.17	0.43	3.79	-0.53	0.2	0.51
	OB2	0.63	-0.76	-1.58	11.91	0.62	3.73	-0.02	1.39	0.67
	OB3	0.59	-0.77	-1.61	11.24	0.59	3.72	-0.51	1.4	0.67
	DP1	1.85	-0.09	1.26	4.39	1.84	3.38	2.24	2.02	1.94
	OB4	1.04	-0.50	-2.73	-1.90	1.02	5.25	-0.47	0.55	1.08
	OB5	0.44	-0.84	-2.02	-0.17	0.43	3.8	-0.82	0.18	0.52
	DP2	1.97	-0.08	1.71	15.26	1.94	3.06	2.22	2.98	2.01
	OB6	1.18	-0.63	-2.53	-1.70	1.17	5.15	-0.12	0.68	1.21
	OB7	1.06	-0.71	-2.25	8.35	1.05	5.08	0.31	1.31	1.14
	OB8	1.05	-0.73	-2.45	2.4	1.04	5.24	-0.03	0.88	1.08
	DP3	3.28	0.06	4.34	19.74	3.26	3.05	3.70	4.58	3.34
	DP4	3.19	-0.21	3.67	6.6	3.2	3.38	3.90	3.52	3.21
	DP5	3.09	-0.26	3.52	16.15	3.07	3.06	3.64	4.26	3.18
	DP6	3.52	0.11	3.96	16.96	3.50	3.52	4.38	4.38	3.55
	OB9	1.52	-0.96	-0.95	17.88	1.51	5.21	0.55	2.72	1.57
	DP7	3.37	-0.22	4.37	18.8	3.35	2.93	4.18	4.74	3.46
	DP8	3.33	-0.29	3.71	16.52	3.30	3.35	4.12	4.51	3.41
	OB10	0.51	-0.75	-1.93	-0.12	0.50	3.65	-0.58	0.21	0.57
	OB11	1.99	-0.60	-1.73	8.84	1.98	5.79	0.79	2.49	2.05
	DP10	3.34	-0.27	3.73	16.56	3.30	3.44	4.05	4.52	3.38
	DP11	3.98	0.04	4.41	15.53	4.00	4.23	4.82	4.90	4.05
	C_1	4.28	-0.56	3.79	20.62	4.28	6.44	4.55	5.39	4.35
	C_2	4.23	-0.95	3.42	14.45	4.18	6.19	4.31	4.95	4.33
	C_3	4.53	-0.83	3.80	17.72	4.53	6.34	4.82	5.47	4.60

Table S10: Continuation of Table S8.

		Supersystem	No embedding				Mulliken embedding			
N	Name	Energy	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)
3	C_1	0.00	0.00	0.00	N/A	N/A	0.00	0.00	N/A	N/A
	C_3	0.86	-0.02	0.77	N/A	N/A	0.32	0.83	N/A	N/A
	C_{3h}	1.65	-0.04	1.03	N/A	N/A	1.27	1.48	N/A	N/A
4	S_4	0.00	0.00	0.00	0.00	N/A	0.00	0.00	0.00	N/A
	C_i	1.01	-0.03	0.81	1.24	N/A	1.01	0.82	0.99	N/A
	C_4	2.41	-0.10	1.97	2.47	N/A	1.82	2.23	2.4	N/A
	C_{4h}	3.27	-0.22	2.03	3.42	N/A	2.95	3.07	3.39	N/A
5	C_1	0.00	0.00	0.00	0.00	N/A	0.00	0.00	0.00	N/A
	C_5	3.13	-0.18	2.05	3.26	N/A	2.65	2.79	3.00	N/A
	C_{5h}	3.70	-0.33	2.24	3.20	N/A	3.24	3.19	3.58	N/A
6	Prism	1.75	0.21	-3.70	-2.56	N/A	1.83	0.79	1.25	N/A
	Cyclic	0.00	0.00	0.00	0.00	N/A	0.00	0.00	0.00	N/A
	Cage	1.55	0.00	-2.94	-3.36	N/A	1.21	0.69	1.05	N/A
	Book1	0.37	0.01	-2.16	-2.41	N/A	0.39	0.06	0.05	N/A
	Book2	0.69	0.05	-2.31	-1.31	N/A	0.57	0.37	0.30	N/A
	Bag	1.63	0.20	-0.75	-0.23	N/A	1.53	1.38	1.47	N/A
	Boat1	1.32	-0.04	0.42	1.44	N/A	1.37	1.23	1.19	N/A
	Boat2	1.37	-0.06	0.35	0.82	N/A	1.75	1.32	1.30	N/A

Table S11: Relative energies (kcal/mol) of $(H_2O)_N$ clusters at the B3LYP/6-31+G(d,2p) level of theory. Cluster geometries and nomenclature are taken from Ref. 7.

N	Name	Supersystem Energy	No embedding				Mulliken embedding			
			SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)
7	OptA	0.00	0.00	0.00	0.00	N/A	0.00	0.00	0.00	N/A
	OptB	1.24	-0.05	1.93	-0.10	N/A	1.39	1.33	N/A	N/A
	OptC	1.59	-0.15	4.19	3.30	N/A	2.41	2.11	1.94	N/A
	OptD	2.35	-0.29	3.86	-1.33	N/A	3.76	2.36	2.21	N/A
	PR2	0.41	0.12	0.65	-0.05	N/A	0.50	0.42	0.44	N/A
	PR3	0.6	0.01	0.43	0.08	N/A	0.74	0.57	0.78	N/A
	CA1	1.10	-0.11	1.82	1.03	N/A	1.46	1.17	1.31	N/A
	CA2	1.96	0	2.89	1.30	N/A	2.30	2.24	2.17	N/A
	CH1	1.47	-0.14	4.07	2.68	N/A	2.54	2.06	1.83	N/A
	CH2	0.98	-0.13	3.63	3.39	N/A	1.89	1.54	1.30	N/A
	CH3	1.47	-0.14	4.07	2.67	N/A	2.54	1.99	1.83	N/A
8	D_{2d}	0.20	-0.04	-0.26	-10.49	N/A	0.26	-0.03	-0.68	N/A
	S_4	0.00	0.00	0.00	0.00	N/A	0.00	0.00	0.00	N/A
	C_i	3.13	0.06	2.83	4.40	N/A	4.00	2.99	3.15	N/A
	C_s	4.21	0.41	3.40	5.14	N/A	5.24	3.75	3.89	N/A
	C_2	3.06	0.00	2.48	-4.61	N/A	4.34	2.89	2.25	N/A
	C_1b	4.11	-0.01	3.01	5.69	N/A	5.71	3.62	2.64	N/A
	C_1a	4.36	-0.23	2.69	-4.61	N/A	6.23	3.89	3.48	N/A
	C_1 c	4.51	-0.18	2.96	4.66	N/A	5.91	3.92	4.30	N/A
	Noncubic	4.79	-0.38	4.02	-1.90	N/A	6.69	4.43	4.35	N/A
9	D_{2d}Dh	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	S_4Dah 1	0.41	0.12	0.75	11.05	0.56	0.1	0.43	1.08	0.44
	S_4Dah 2	0.40	0.13	0.77	11.03	0.34	0.11	0.36	1.16	0.43
	S_4DDh 1	0.31	0.03	0.24	9.56	0.55	0.42	0.41	1.03	0.36
	S_4DDh 2	0.41	0.02	0.11	9.51	-0.06	0.33	0.28	0.95	0.30
	D_{2d}Dah	0.7	0.08	0.54	0.39	-0.41	1	0.6	0.44	0.74
	S_4Dah1	0.67	0.11	0.55	8.07	1.21	0.69	0.64	1.02	0.71
	S_4Dah2	0.87	0.1	0.75	7.47	1.21	1.04	0.74	1.35	0.94

Table S12: Continuation of Table S11.

		Supersystem	No embedding				Mulliken embedding			
N	Name	Energy	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)
10	PP1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	PP2	0.15	0.04	-0.04	16.26	1.23	0.10	0.19	1.58	0.13
	PP3	0.87	0.09	-0.30	3.84	2.76	1.18	0.78	0.87	0.78
	PP4	0.80	0.16	0.22	13.72	1.42	0.87	1.02	1.82	0.88
	PP5	0.88	0.01	0.15	8.43	2.11	1.12	0.99	1.49	0.96
	OB1	2.75	-0.55	0.77	1.84	3.70	3.34	2.09	2.48	2.77
	OB2	2.79	-0.53	0.66	12.56	3.37	3.11	2.43	3.37	2.84
	OB3	2.74	-0.53	0.68	12.29	3.30	3.13	1.96	3.39	2.84
	DP1	2.15	0.04	1.66	3.58	3.19	3.16	2.32	2.27	2.11
	OB4	3.05	-0.31	0.38	-0.29	6.00	4.55	1.94	2.61	3.08
	OB5	2.74	-0.55	0.76	1.84	3.69	3.35	1.78	2.43	2.76
	DP2	2.18	0.03	1.86	13.63	3.27	2.76	2.13	3.07	2.31
	OB6	3.27	-0.41	0.53	-0.08	5.31	4.26	2.37	2.74	3.33
	OB7	3.1	-0.46	0.91	9.41	5.40	4.02	2.63	3.23	3.19
	OB8	3.16	-0.47	0.98	3.79	6.17	4.27	2.34	2.93	3.18
	DP3	2.22	0.06	2.88	16.09	5.39	2.66	2.10	3.25	2.32
	DP4	2.44	-0.10	2.77	4.67	6.98	3.22	2.63	2.67	2.52
	DP5	2.31	-0.14	2.39	13.64	5.21	3.12	2.44	3.28	2.41
	DP6	2.28	0.09	2.83	13.39	6.56	2.79	2.73	2.88	2.34
	OB9	3.58	-0.71	1.53	18.12	3.9	4.01	2.85	4.43	3.69
	DP7	2.25	-0.08	2.5	15.92	4.14	2.63	2.47	3.29	2.32
	DP8	2.59	-0.16	2.33	13.76	5.36	3.33	2.90	3.56	2.67
	OB10	2.73	-0.50	0.84	1.72	3.93	3.20	1.98	2.45	2.77
	OB11	3.90	-0.36	0.90	9.55	5.1	4.65	3.06	4.23	3.96
	DP10	2.63	-0.15	2.39	13.77	5.35	3.39	2.85	3.6	2.74
	DP11	2.93	0.08	3.81	12.53	7.07	3.73	3.41	3.67	3.03
	C_1	4.63	-0.33	4.37	18.4	6.62	5.74	4.59	5.49	4.70
	C_2	4.71	-0.52	4.05	13.21	8.30	5.83	4.37	5.46	4.78
	C_3	4.74	-0.42	3.90	15.76	7.41	5.97	4.51	5.56	4.82

Table S13: Continuation of Table S11.

Cluster	Supersystem Energy	No embedding				Mulliken embedding			
		SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)
1	-859.762827	-859.555124	-859.792072	-859.741326	-859.761843	-859.833982	-859.761556	-859.786576	-859.762588
2	-859.777894	-859.570196	-859.803077	-859.764248	-859.777709	-859.861604	-859.776333	-859.790504	-859.777830
3	-859.749972	-859.539163	-859.770032	-859.756475	-859.749724	-859.841423	-859.748281	-859.765854	-859.749886
4	-859.756497	-859.547927	-859.780230	-859.751765	-859.756203	-859.832971	-859.754989	-859.778415	-859.756301
5	-859.733434	-859.535780	-859.757520	-859.722162	-859.732399	-859.813402	-859.731829	-859.749879	-859.733194
6	-859.748323	-859.542489	-859.777686	-859.748767	-859.748000	-859.843935	-859.746906	-859.764451	-859.748196
7	-859.759586	-859.557143	-859.784282	-859.743196	-859.757903	-859.836458	-859.757569	-859.770988	-859.759368
8	-859.759179	-859.561192	-859.783586	-859.766006	-859.758626	-859.841430	-859.757161	-859.773615	-859.759013
9	-859.734403	-859.539272	-859.760121	-859.707940	-859.732044	-859.818888	-859.731870	-859.757653	-859.734059
10	-859.738071	-859.524482	-859.766504	-859.715685	-859.738548	-859.822728	-859.736744	-859.748054	-859.737784

Table S14: Energies (a.u.) of ten $F^-(H_2O)_{10}$ clusters at the HF/6-31+G* level.

	Fixed ChElPG				XPol ChElPG			
Cluster	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)
1	-859.834059	-859.761421	-859.782159	-859.762646	-859.823359	-859.749320	-859.772888	-859.765494
2	-859.861201	-859.776353	-859.787690	-859.777853	-859.851568	-859.761565	-859.780950	-859.779457
3	-859.840063	-859.748738	-859.762484	-859.749903	-859.833552	-859.734287	-859.757869	-859.749325
4	-859.832121	-859.755276	-859.776480	-859.756313	-859.821756	-859.748734	-859.760049	-859.756369
5	-859.812575	-859.731281	-859.746539	-859.733237	-859.806611	-859.719203	-859.737586	-859.732098
6	-859.843304	-859.746666	-859.758017	-859.748263	-859.834205	-859.733643	-859.751395	-859.747965
7	-859.836091	-859.757464	-859.767353	-859.759426	-859.830303	-859.743896	-859.764757	-859.757346
8	-859.840893	-859.757127	-859.772069	-859.759023	-859.834160	-859.742565	-859.763559	-859.758746
9	-859.818297	-859.731516	-859.753166	-859.734157	-859.812264	-859.718236	-859.747331	-859.735429
10	-859.822182	-859.736867	-859.747303	-859.737796	-859.816365	-859.722666	-859.743925	-859.737777

Table S15: Energies (a.u.) of ten $F^-(H_2O)_{10}$ clusters at the HF/6-31+G* level using ChElPG charges.

Cluster	Supersystem Energy	No embedding				Mulliken embedding			
		SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)
1	-864.406751	-864.194045	-864.448849	-864.386136	-864.404565	-864.450022	-864.408762	-864.424608	-864.406858
2	-864.420824	-864.206499	-864.459011	-864.408707	-864.420959	-864.472568	-864.422746	-864.430798	-864.420927
3	-864.402009	-864.183337	-864.433086	-864.408080	-864.401782	-864.458303	-864.402487	-864.414977	-864.402206
4	-864.408853	-864.194348	-864.443606	-864.403346	-864.408652	-864.455941	-864.410430	-864.426974	-864.408485
5	-864.387366	-864.180978	-864.422960	-864.376987	-864.385067	-864.434202	-864.387933	-864.400337	-864.387432
6	-864.403808	-864.188792	-864.446841	-864.403292	-864.402534	-864.463443	-864.405864	-864.415634	-864.403631
7	-864.403109	-864.195477	-864.438822	-864.388964	-864.399967	-864.451239	-864.404122	-864.412640	-864.402512
8	-864.401550	-864.198217	-864.444970	-864.408187	-864.400644	-864.455279	-864.402606	-864.413759	-864.401508
9	-864.386238	-864.184090	-864.427676	-864.359608	-864.382326	-864.438811	-864.388983	-864.403103	-864.386301
10	-864.392280	-864.173995	-864.433627	-864.370832	-864.392323	-864.445070	-864.395731	-864.398870	-864.392424

Table S16: Energies (a.u.) of ten $F^-(H_2O)_{10}$ clusters at the B3LYP/6-31+G(d,2p) level.

	Fixed ChElPG				XPol ChElPG			
Cluster	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)
1			-864.436396	-864.389485			-864.409151	-864.409285
2			-864.458991	-864.401111			-864.419565	-864.421441
3			-864.446140	-864.382120			-864.405784	-864.399333
4			-864.440626	-864.398192			-864.407167	-864.407086
5			-864.423902	-864.372187			-864.386103	-864.385176
6			-864.450556	-864.385739			-864.400814	-864.401722
7			-864.441674	-864.383986			-864.405056	-864.398665
8			-864.444330	-864.382378			-864.402397	-864.399636
9			-864.428724	-864.369373			-864.390867	-864.385829
10			-864.434978	-864.373826			-864.391878	-864.390512

Table S17: Energies (a.u.) of ten $F^-(H_2O)_{10}$ clusters at the B3LYP/6-31+G(d,2p) level using ChElPG charges.

	Supersystem	No embedding				Mulliken embedding			
Cluster	Energy	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)
1	9.45	9.46	6.91	14.38	9.96	17.33	9.27	2.46	9.56
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	17.52	19.47	20.74	4.88	17.56	12.66	17.6	15.47	17.54
4	13.43	13.97	14.34	7.83	13.50	17.97	13.39	7.59	13.51
5	27.90	21.60	28.59	26.41	28.43	30.25	27.93	25.49	28.01
6	18.56	17.39	15.93	9.71	18.64	11.09	18.47	16.35	18.60
7	11.49	8.19	11.79	13.21	12.43	15.78	11.77	12.25	11.59
8	11.74	5.65	12.23	-1.10	11.97	12.66	12.03	10.6	11.8
9	27.29	19.41	26.96	35.33	28.66	26.8	27.9	20.61	27.47
10	24.99	28.69	22.95	30.47	24.57	24.40	24.84	26.64	25.13

Table S18: Relative energies (kcal/mol) of ten $F^-(H_2O)_{10}$ clusters at the HF/6-31+G* level.

	Fixed ChElPG				XPol ChElPG			
Cluster	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)
1	17.03	9.37	3.47	9.54	17.70	7.68	5.06	8.76
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	13.26	17.33	15.82	17.54	11.31	17.12	14.48	18.91
4	18.25	13.23	7.03	13.52	18.71	8.05	13.12	14.49
5	30.51	28.28	25.82	28.00	28.21	26.58	27.21	29.72
6	11.23	18.63	18.62	18.57	10.90	17.52	18.55	19.76
7	15.76	11.85	12.76	11.56	13.34	11.09	10.16	13.87
8	12.74	12.06	9.80	11.82	10.92	11.92	10.91	13.00
9	26.92	28.14	21.66	27.42	24.66	27.19	21.10	27.63
10	24.48	24.78	25.34	25.14	22.09	24.41	23.23	26.15

Table S19: Relative energies (kcal/mol) of ten $F^-(H_2O)_{10}$ clusters at the HF/6-31+G* level using ChElPG charges.

Cluster	Supersystem Energy	No embedding				Mulliken embedding			
		SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2))
1	8.83	7.81	6.38	14.16	10.29	14.15	8.78	3.88	8.83
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	11.81	14.53	16.27	0.39	12.03	8.95	12.71	9.93	11.75
4	7.51	7.62	9.67	3.36	7.72	10.43	7.73	2.40	7.81
5	21.00	16.01	22.62	19.90	22.52	24.08	21.85	19.11	21.02
6	10.68	11.11	7.64	3.40	11.56	5.73	10.59	9.52	10.85
7	11.12	6.92	12.67	12.39	13.17	13.38	11.69	11.39	11.56
8	12.09	5.20	8.81	0.33	12.75	10.85	12.64	10.69	12.19
9	21.70	14.06	19.66	30.81	24.24	21.18	21.19	17.38	21.73
10	17.91	20.40	15.93	23.77	17.97	17.26	16.95	20.04	17.89

Table S20: Relative energies (kcal/mol) of ten $F^-(H_2O)_{10}$ clusters at the B3LYP/6-31+G(d,2p) level.

Cluster	Fixed ChElPG				XPol ChElPG			
	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)
1			14.18	7.30			6.53	7.63
2			0.00	0.00			0.00	0.00
3			8.06	11.92			8.65	13.87
4			11.52	1.83			7.78	9.01
5			22.02	18.15			21.0	22.76
6			5.29	9.65			11.77	12.37
7			10.87	10.75			9.10	14.29
8			9.20	11.76			10.77	13.68
9			18.99	19.92			18.01	22.35
10			15.07	17.12			17.37	19.41

Table S21: Relative energies (kcal/mol) of ten $F^-(H_2O)_{10}$ clusters at the B3LYP/6-31+G(d,2p) level using ChElPG charges.

	Supersystem	No embedding				Mulliken embedding			
Cluster	Energy	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)
1	-4333.270624	-4332.461087	-4332.808846	-4332.897030	-4333.036125	-4333.45381	-4333.243934	-4333.266474	-4333.272388
2	-4333.186038	-4332.415237	-4332.736489	-4332.841170	-4332.969250	-4333.367291	-4333.137009	-4333.160769	-4333.169989
3	-4333.134450	-4332.402889	-4332.722667	-4332.834066	-4332.924915	-4333.330114	-4333.095429	-4333.117783	-4333.122337
4	-4333.162441	-4332.405300	-4332.682170	-4332.765592	-4332.831698	-4333.358751	-4333.118920	-4333.139055	-4333.152386
5	-4333.061269	-4332.361730	-4332.654411	-4332.752520	-4332.832271	-4333.238178	-4333.016137	-4333.031523	-4333.040740
1	-4357.418890	-4356.760109	-4357.069920	-4357.128588	-4357.241931	-4357.692411	-4357.471720	-4357.473041	-4357.459123
2	-4357.372406	-4356.738793	-4357.031668	-4357.107017	-4357.207536	-4357.632285	-4357.403044	-4357.406979	-4357.394373
3	-4357.329735	-4356.723424	-4357.014231	-4357.098449	-4357.172257	-4357.592374	-4357.361195	-4357.365351	-4357.355284
4	-4357.354852	-4356.717836	-4356.976394	-4357.041913	-4357.092634	-4357.611741	-4357.385923	-4357.394273	-4357.393013
5	-4357.276181	-4356.687929	-4356.962806	-4357.034294	-4357.096930	-4357.508767	-4357.293474	-4357.295458	-4357.293070

Table S22: Top: Energies (a.u.) of $(H_2O)_{57}$ clusters at the HF/6-31G* level of theory. Bottom: Energies (a.u.) of $(H_2O)_{57}$ at the B3LYP/6-31+G(d,2p) level of theory.

	No embedding				Mulliken embedding			
Cluster	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)	SFM1(1)	SFM1(2)	GEBF(1)	GEBF(2)
1	8.91	5.08	4.11	2.58	-2.02	0.29	0.05	-0.02
2	8.49	4.95	3.80	2.39	-2.00	0.54	0.28	0.18
3	8.05	4.53	3.31	2.31	-2.15	0.43	0.18	0.13
4	8.34	5.29	4.37	3.64	-2.16	0.48	0.26	0.11
5	7.70	4.48	3.40	2.52	-1.95	0.50	0.33	0.23
1	7.25	3.84	3.20	1.95	-3.01	-0.58	-0.60	-0.44
2	6.98	3.75	2.92	1.82	-2.86	-0.34	-0.38	-0.24
3	6.67	3.47	2.55	1.73	-2.89	-0.35	-0.39	-0.28
4	7.01	4.17	3.45	2.89	-2.83	-0.34	-0.43	-0.42
5	6.48	3.45	2.66	1.97	-2.56	-0.19	-0.21	-0.19

Table S23: Top: Signed Errors (kcal/mol/monomer) of $(H_2O)_{57}$ at the HF/6-31G* level of theory. Bottom: Signed Errors (kcal/mol/monomer) of $(H_2O)_{57}$ at the B3LYP/6-31+G(d,2p) level of theory.