1-1-2014

In silico enzyme modelling

Xiao Zhu
Texas Advanced Computing Centre

Yang Yang
Rowan University, yy922@uowmail.edu.au

Haibo Yu
University of Wollongong, hyu@uow.edu.au

In silico enzyme modelling

Abstract

The 2013 Nobel Prize in Chemistry went to Martin Karplus, Michael Levitt and Arieh Warshel for their pioneering work on computer modelling, specifically, the 'development of multiscale models of complex chemical systems' (1). This award not only recognises the critical contributions by the three laureates to the field of molecular simulations, but also underscores the broad impact that computer simulations have made in fields as diverse as chemistry, biophysics, enzymology and material sciences. This review will present an overview of computational enzymology, a rapidly maturing field where multiscale modelling plays a key role in deciphering enzymatic catalysis (2-4).

Disciplines

Medicine and Health Sciences | Social and Behavioral Sciences

Publication Details

Zhu, X., Yang, Y. & Yu, H. (2014). In silico enzyme modelling. Australian Biochemist, 45 (2), 12-15.

In Silico *Enzyme Modelling*

Xiao Zhu[1], Yang Yang[2] and Haibo Yu[3]*

[1]*High Performance Computing Group, Texas Advanced Computing Centre, Austin, Texas 78758, USA*
[2]*Department of Chemistry and Biochemistry, Rowan University, Glassboro, New Jersey 08028, USA*
[3]*School of Chemistry, University of Wollongong, NSW 2522*
**Corresponding author: hyu@uow.edu.au*

The 2013 Nobel Prize in Chemistry went to Martin Karplus, Michael Levitt and Arieh Warshel for their pioneering work on computer modelling, specifically, the 'development of multiscale models of complex chemical systems' (1). This award not only recognises the critical contributions by the three laureates to the field of molecular simulations, but also underscores the broad impact that computer simulations have made in fields as diverse as chemistry, biophysics, enzymology and material sciences. This review will present an overview of computational enzymology, a rapidly maturing field where multiscale modelling plays a key role in deciphering enzymatic catalysis (2–4).

Enzymes are superb catalysts in nature; the enormous rate accelerations of enzymatic reactions over their uncatalysed counterparts typically range between 10^{10} to 10^{20} fold (5). Unraveling how enzymes 'work' – how they speed up difficult chemical transformations with high efficiency and specificity, under mild physiological conditions, is a question of fundamental importance in biochemistry. At the same time, knowledge of the origin of this catalytic power on a molecular level has many practical applications in the form of novel artificial catalysts and enzyme inhibitors that might act as drugs. Over the past decades, enormous progress has been made in understanding the molecular basis of enzymatic reactions and their underlying mechanisms. Experimentally, this has been achieved through sustained efforts in structural biology, mutagenesis and enzyme kinetics. Such achievements have led to the *de novo* design of novel enzymes (6) and numerous inhibitors for use as drugs (7).

Why is QM/MM Used to Model Enzymatic Reactions?

Because of the complexity of enzymes, potential ambiguity in the interpretation, and limitation in the spatial and temporal resolution of many experimental observables, a number of key mechanistic questions remain to be answered. For instance, it has been debated whether linear free energy relations and kinetic isotopic effects can unambiguously define the nature of transition states (8). As a result, a complete and quantitative understanding of enzyme catalysis is still lacking; this is evident from our inability to engineer efficient catalysts that match naturally evolved enzymes, in addition to the many mechanistic puzzles uncovered over the years. The seminal work by Karplus, Levitt and Warshel demonstrated that well-calibrated, combined quantum mechanics/molecular mechanics (QM/MM) computational methods are capable of producing reliable predictions for structural and energetic properties and are tremendously valuable for detailed, atomic-level analyses of underlying mechanisms (9,10). The basic strategy in the QM/MM approach is straightforward (**Fig 1a**): the reactive part, a small portion of the system, is treated quantum mechanically; popular choices include fast, approximate methods (11) (e.g. SCC-DFTB, DFTB3, OMX) and more accurate, but computationally more expensive density functional theory or *ab initio* methods (12). The QM treatment allows the modelling of the electronic rearrangements involved in bond breakage and formation during a chemical reaction. On the other hand, the large non-reactive part of the enzyme and surrounding solvent are described more simply by empirical molecular mechanics (MM), the so-called force field (**Fig 1b**). In such a force field, the intra-molecular and inter-molecular interactions are represented by simple mathematical functions. The interactions between the QM and MM regions are included *via* different coupling schemes. Combined QM/MM simulations provide a direct window to monitor the electronic, structural and dynamical properties of an enzymatic reaction as it occurs (**Fig 1c** and **1d**).

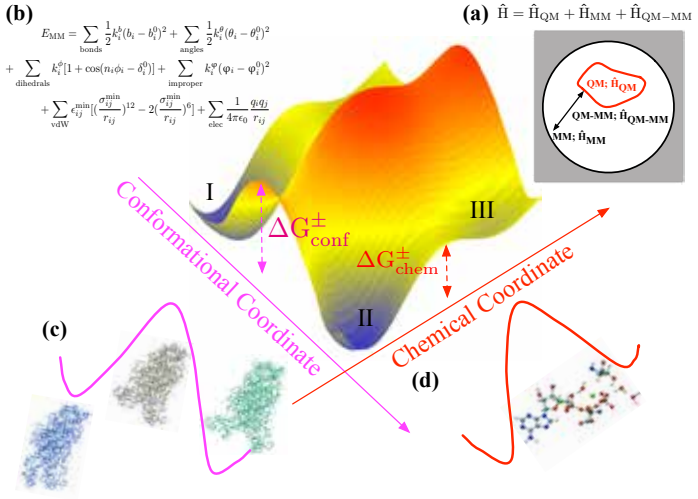


Fig. 1. The major components of computational enzymology include:
(a) combined QM/MM methods;
(b) classical force fields;
(c) multiscale modelling approaches to explore the free energy landscape along the conformational coordinate; and
(d) methods to calculate accurate energetics along the chemical coordinate. States I/II represent two conformational states along the conformational coordinate; while States II/III are the reactant state and the product state along one possible reaction pathway defined by the chemical coordinate.

What Can QM/MM Offer Biochemists?

Over the years, biomolecular simulations have become an indispensable tool for studying the structural, dynamic and functional aspects of biomolecular systems in atomic detail and are now widely used to interpret experimental data, test mechanistic hypotheses, and inspire new experiments (13). Since the very first study by Warshel and Levitt on hen egg white lysozyme in 1976 (10), QM/MM simulations of enzymatic reactions have contributed significantly to the field of enzymology. This has been made possible by increased computing power, greater accessibility of simulation software and developments in methodology.

Formulating and testing mechanistic hypotheses: Determination of enzyme mechanisms has proven difficult in many cases, since differentiating between alternative possibilities in experiments is often challenging; even the common examples of enzyme mechanisms described in classical biochemistry textbooks have been challenged by recent investigations (14,15). Computational studies can be carried out on the wildtype enzymes and substrates, thereby avoiding the possible confusion and controversies in the literature where experimental modifications of enzymes or/and substrates are required to probe the wildtype mechanisms. From combined QM/MM simulations, the reaction thermodynamics and kinetics can be evaluated for different reaction pathways, and thus validate or disapprove them (16). One example to show the value of such QM/MM simulations for clarifying catalytic mechanisms is recent work on ATP hydrolysis catalyzed by myosin (shown in **Fig. 2**) (17). In the hydrolysis-competent state of myosin, it is generally believed that ATP hydrolysis operates via an associative mechanism, with the conserved Ser236 serving as the proton relay (**Fig. 2a**). The rate-limiting step is the formation of the 'pentavalent' intermediate state (**Fig. 2b** and **2c**). The calculated energy barrier is ~16 kcal/mol, in agreement with experimentally estimated values of 15–17 kcal/mol. We provide structural and energetic evidence to support the idea that regulation of ATP hydrolysis activity is not limited to residues in the immediate environment of ATP. Efficient hydrolysis of ATP depends not only on the proper orientation of the lytic water but also on the structural stability of several nearby residues in the active site of myosin. More importantly, our results clearly indicate that turning on the ATPase activity requires not only structural displacement close to the active site but also structural transitions beyond the immediate environment of ATP, which have been proposed previously to be ultimately coupled to the rotation of the converter subdomain 40 Å away (schematically shown in **Fig. 1c**). Additionally, such QM/MM simulations also offer an efficient way to predict the effects of mutagenesis by examining the contributions from a specific residue to the energetic properties and this in turn contributes to the practical application of enzymology.

Characterisation of the key states along the catalytic pathways: Combined QM/MM simulations can also provide structural and electronic information on the species appearing along the reaction pathway: from the Michaelis complex, to the transition state and the intermediate state, and finally to the product state. This knowledge is of great value for mechanism-based inhibitor design, which has been exemplified by the success in developing inhibitors as drugs. Most obviously, simulations at atomistic scale are capable of capturing the transition state structure and their key interactions that are not directly accessible by experiments. Enzymologists have theorised for many years that enzymes are able to recognise and stabilise the transition states, and that this stabilisation is at the heart of enzyme catalysis. For instance, a recent study showed that the alkaline phosphatase superfamily can recognise and stabilise different types of transition states in the same active site (18). Detailed QM/MM analyses provide a molecular rationale for the catalytic promiscuity these enzymes exhibit towards a broad class of substrates. Another feature noted about enzymatic catalysis is that despite the limited functional groups available, enzymes can catalyse the

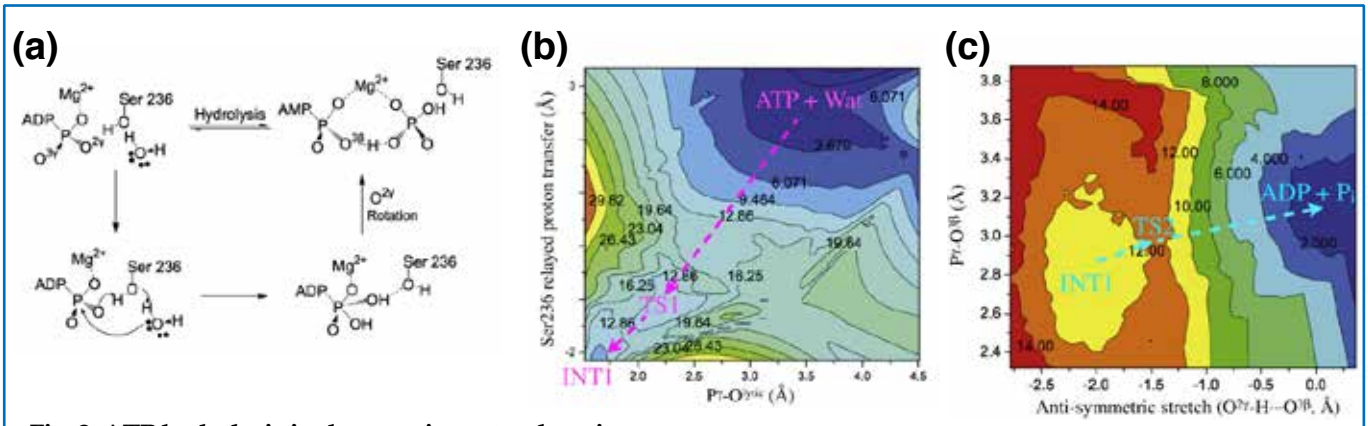
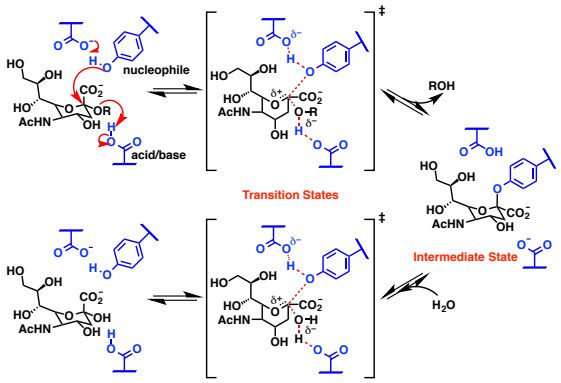


Fig. 2. ATP hydrolysis in the myosin motor domain.
(a) The schematics of the reaction pathway followed in the QM/MM simulations, which involves Ser236 as the proton relay. **(b)** The two-dimensional potential of mean force (kcal/mol) for the first step of ATP hydrolysis in the pre-powerstroke state; the x-axis is the distance between the lytic water oxygen and Pγ of ATP, and the y-axis is a collective coordinate that describes the relayed proton transfer involving the lytic water, Ser236, and the γ phosphate of ATP. **(c)** The two-dimensional potential of mean force for the second step of ATP hydrolysis in the pre-powerstroke state; the x-axis is the antisymmetric stretch involving the O2γ group and O3β of ATP, and the y-axis is the Pγ-O3β distance. Adapted from ref. 17 with permission.

Table 1. p*K*a for the general acid/base Asp151 in influenza virus neuraminidase (adapted from Ref. 20). Three different approaches were used to calculate p*K*a. **PBEQ:** p*K*a calculations based on PBEQ; **PROPKA:** p*K*a calculations based on PROPKA; **QM/MM FEP:** p*K*a calculations based on DFTB3/MM FEP with explicit representation of solvent; **Prot. States:** the protonation states required to perform the proposed catalytic roles.

State	PBEQ	PROPKA	QM/MM	Prot. States
apo	4.1	3.6	4.1	Deprot.
holo	5.1	3.6	8.7	Prot.
int.	4.4	4.2	6.3	Deprot.



myriad reactions necessary to sustain life. This has been attributed to their ability to modulate the protonation states of the catalytic residues so that they can serve either as nucleophilic, electrophilic, or general acid–base catalysts (19). Recently, we have shown through pKa calculations based on combined QM/MM simulations that the general acid/base in trans-sialidase and neuraminidase can go through the so-called 'pKa cycling' process to achieve the proper protonation states required for their roles (20). For the first time, we provide a structural and energetic basis for the proposed pKa cycling in these enzymes.

What Does QM/MM Hold for the Future?

In 2002, Hansson *et al.* summarised the future of biomolecular simulations in general as "bigger, better, faster" (21), which we believe nicely depicts the future of computational enzymology.

Bigger: Like all branches of computational science, computational enzymology benefits from the seemingly never-ending improvements in computer hardware. This will allow us to study chemical reactions in much larger systems with a more realistic representation of the molecular environment. QM simulation of a small protein with explicit solvent for 350 ps has been reported (22). Many enzymes are able to couple the chemical step in the active site with large-scale conformational transitions that occur in parts of the enzyme far from the active site. The unsolved issues surrounding the complexity and the multiscale nature of the coupling between the chemical step and structural transition are the subjects of many heated debates (as illustrated in **Fig. 1**). For instance, it has been debated in the literature whether enzyme dynamics contributes significantly to catalysis. It is expected that with the development of multiscale modelling techniques in computational enzymology, simulation will play an important role in addressing these challenging questions (23,24).

Better: The accuracy of the adopted QM/MM model largely determines the reliability of computational studies (25). Sustained efforts have been directed towards improving the various components of combined QM/MM simulations (**Fig. 1**). Semi-empirical models have been one of the most attractive choices for the QM model in the combined QM/MM simulations, due to their efficiency and ability to carry out free energy simulations where entropic effects are naturally taken into account (11,26). With the need for the power to make predictions with near chemical accuracy (i.e. within 1 kcal/mol) and the realisation of the possible limitations of the available semi-empirical models, methodology development is currently focused in two directions. First, substantial efforts are being devoted to improving the accuracy of the semi-empirical models (27,28). Second, algorithms are being designed to obtain, with practical feasibility, reaction barriers and energies at the *ab initio* level (12,29). From the perspective of the adopted MM potentials, polarisable force fields are currently an area under active development (30). Incorporating explicit polarisation will provide a better description of the changing electrostatics that accompany the chemical reactions (31). It is also critical to compute the long-range electrostatic interactions accurately in the QM/MM simulations as they have a profound effect on structural, dynamic and energetic properties (32,33).

Faster: Adequate sampling is also crucial for reliable results when modelling enzymatic reactions. It has been shown that the extensive sampling of the configuration space of both the reactive part and the enzyme are essential for properly determining the energetics (26,34). The reaction energies and activation energies based on energy minimisation studies heavily depend on the exact protein structure used in the simulation. With the development of computing hardware, such problems can be partially alleviated by carrying out sufficiently long molecular dynamics simulations to sample multiple, energetically relevant configurations. Novel techniques that enhance the sampling of rare events offer an alternative approach to overcoming the problems associated with insufficient sampling (35). This is particularly attractive for a QM/MM simulation when considering the fact that the amount of conformational sampling is often limited by its high computational cost. Moreover, enzymatic reactions are likely coupled to significant protein conformational changes and solvent responses, some of which are not easily captured with brute-force simulations due to the existence of significant barriers between different states.

Considering the complexity of enzymatic reactions and the possible limitations of the experimental and computational techniques, the most productive avenue is to use an integrated computational/experimental framework, designed to maximise the complementarity between computation and experiment. Indeed, synergistic collaborations between experimental and computational

enzymologists are common, and are poving to be extremely beneficial in addressing mechanistic puzzles that cannot otherwise be conclusively resolved. With the development of computational methods and the ever-increasing availability of computing resources, the field of computational enzymology will continue to flourish and play an increasingly important role in various aspects of biochemistry and medicinal chemistry.

References

1. The Nobel Prize in Chemistry 2013, Press Release, Nobel Media AB
2. Lin, H., and Truhlar, D.G. (2007) *Theor. Chem. Acc.* 117, 185-199
3. Senn, H.M., and Thiel, W. (2009) *Angew. Chem. Int. Ed. Engl.* **48**, 1198-1229
4. Lonsdale, R., Ranaghan, K.E., and Mulholland, A.J. (2010) *Chem. Commun.* **46**, 2354-2372
5. Wolfenden, R. (2011) *Annu. Rev. Biochem.* **80**, 645-667
6. Röthlisberger, D., Khersonsky, O., Wollacott, A.M., *et al.* (2008) *Nature* **453**, 190-195
7. Imming, P., Sinning, C., and Meyer, A. (2006) *Nat. Rev. Drug Discov.* **5**, 821-834
8. Aqvist, J., Kolmodin, K., Florian, J., and Warshel, A. (1999) *Chem. Biol.* **6**, R71-R80
9. Warshel, A., and Karplus, M. (1972) *J. Am. Chem. Soc.* **94**, 5612-5625
10. Warshel, A., and Levitt, M. (1976) *J. Mol. Biol.* **103**, 227-249
11. Thiel, W. (2014) *WIREs Comput. Mol. Sci.* **4**, 145-157
12. Claeyssens, F., Harvey, J.N., Manby, F.R., Mata, R.A., Mulholland, A.J., Ranaghan, K.E., Schütz, M., Thiel, S., Thiel, W., and Werner, H.-J. (2006) *Angew. Chem. Int. Ed. Engl.* **45**, 6856-6859
13. van Gunsteren, W.F., Bakowies, D., Baron, R., *et al.* (2006) *Angew. Chem. Int. Ed. Engl.* **45**, 4064-4092
14. Bowman, A.L. Grant, I.M., and Mulholland, A.J. (2008) *Chem. Commun.* **37**, 4425-4427
15. Vocadlo, D.J., Davies, G.J., Laine, R., and Withers, S.G. (2001) *Nature* **412**, 835-838
16. Warshel, A., Sharma, P.K., Kato, M., Xiang, Y., Liu, H., and Olsson, M.H.M. (2006) *Chem. Rev.* **106**, 3210-3235
17. Yang, Y., Yu, H., and Cui, Q. (2008) *J. Mol. Biol.* **381**, 1407-1420
18. Hou, G., and Cui, Q. (2013) *J. Am. Chem. Soc.* **135**, 10457-10469
19. Harris, T. K., and Turner, G.J. (2002) *IUBMB Life* **53**, 85-98
20. Yu, H., and Griffiths, T.M. (2014) *Phys. Chem. Chem. Phys.* **16**, 5785-5792
21. Hansson, T., Oostenbrink, C., and van Gunsteren, W. (2002) *Curr. Opin. Struct. Biol.* **12**, 190-196
22. Liu, H., Elstner, M., Kaxiras, E., Frauenheim, T., Hermans, J., and Yang, W. (2001) *Proteins* **44**, 484-489
23. Kamerlin, S.C.L., Vicatos, S., Dryga, A., and Warshel, A. (2011) *Annu. Rev. Phys. Chem.* **62**, 41-64
24. Daily, M.D., Yu, H., Phillips, G.N.J., and Cui, Q. (2013) *Top. Curr. Chem.* **337**, 139-164
25. van Gunsteren, W.F., and Mark, A.E. (1998) *J. Chem. Phys.* **108**, 6109-6116
26. Riccardi, D., Schaefer, P., Yang, Y., Yu, H.; Ghosh, N., Prat-Resina, X., Konig, P., Li, G., Xu, D., Guo, H., Elstner, M., and Cui, Q. (2006) *J. Phys. Chem. B* **110**, 6458-6469
27. Yang, Y., Yu, H., York, D., Cui, Q., and Elstner, M. (2007) *J. Phys. Chem. A* **111**, 10861-10873
28. Gaus, M., Cui, Q., and Elstner, M. (2011) *J. Chem. Theory Comput.* **7**, 931-948
29. Kamerlin, S.C., Haranczyk, M., and Warshel, A. (2009) *J. Phys. Chem. B* **113**, 1253-1272
30. Yu, H., and van Gunsteren, W.F. (2005) *Comput. Phys. Commun.* **172**, 69-85
31. Woodcock, H.L., Hodoscek, M., Gilbert, A.T.B., Gill, P.M.W., Schaefer, H.,F., and Brooks, B.R. (2007) *J. Comput. Chem.* **28**, 1485-1502
32. Nam, K., Gao, J., and York, D.M. (2005) *J. Chem. Theory Comput.* **1**, 2-13
33. Schaefer, P., Riccardi, D., and Cui, Q. (2005) *J. Chem. Phys.* **123**, 014905
34. Klähn, M., Braun-Sand, S., Rosta, E., and Warshel, A. (2005) *J. Phys. Chem. B* **109**, 15645-15650
35. Christen, M., and van Gunsteren, W.F. (2008) *J. Comput. Chem.* **29**, 157-166

References from page 18

1. Lukin, J.A., and Ho, C. (2004) *Chem. Rev.* **104**, 1219-1230
2. Perutz, M.F., Rossmann, M.G., Cullis, A.F., Muirhead, H., Will, G., and North, A.C. (1960) *Nature* **185**, 416-422
3. Perutz, M.F., Muirhead, H., Cox, J.M., and Goaman L.C. (1968) *Nature* **219**, 131-139
4. Shibayama, N., Sugiyama, K., Tame, J.R., and Park, S.Y. (2014) *J. Am. Chem. Soc.* **136**, 5097-5105
5. Liu, J., Dong, Y., Zheng, J., He, Y., and Sheng, Q. (2013) *Anal. Sci.* **29**, 1075-1081
6. Kay, L.E. (2005) *J. Magn. Reson.* **173**, 193-207
7. Tilton, R.F. Jr, Dewan, J.C., and Petsko, G.A. (1992) *Biochemistry* **31**, 2469-2481
8. van Gunsteren, W.F., Bakowies, D., Baron, R., Chandrasekhar, I., Christen, M., Daura, X., Gee, P., Geerke, D.P., Glättli, A., Hünenberger, P.H., Kastenholz, M.A., Oostenbrink, C., Schenk, M., Trzesniak, D., van der Vegt, N.F., and Yu, H.B. (2006) *Angew Chem Int Ed Engl* **45**, 4064-4092
9. Ma, J. (2005) *Structure* **13**, 373-380
10. Crystal, R.G. (1989) *Trends Genet.* **5**, 411-417
11. Lomas, D.A., Evans, D.L., Finch, J.T., and Carrell, R.W. (1992) *Nature* **357**, 605-607
12. Kass, I., Knaupp, A.S., Bottomley, S.P., and Buckle, A.M. (2012) *Biophys. J.* **102**, 2856-2865
13. Bu, D.F., Erlander, M.G., Hitz, B.C., Tillakaratne, N.J., Kaufman, D.L., Wagner-McPherson, C.B., Evans, G.A., and Tobin, A.J. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 2115-2119
14. Erlander, M.G., Tillakaratne, N.J., Feldblum, S., Patel, N., and Tobin, A.J. (1991) *Neuron* **7**, 91-100
15. Martin, D.L., and Rimvall, K. (1993) *J. Neurochem.* **60**, 395-407
16. Baekkeskov, S., Aanstoot, H.J., Christgau, S., Reetz, A., Solimena, M., Cascalho, M., Folli, F., Richter-Olesen, H., and De Camilli, P. (1990) *Nature* **347**, 151-156
17. Kass, I., Hoke, D.E., Costa, M.G., *et al.* (2014) *Proc Natl Acad Sci USA* **111**, E2524-E2529